

The Journal of Educational Psychology

*Devoted Primarily to the Scientific Study of Problems of
Learning and Teaching*

BOARD OF EDITORS*

WM. CLARK TROW, *Chairman*
University of Michigan

STEPHEN M. COREY
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

*Appointed by the American Psychological Association

EDITORIAL CONSULTANTS

PAUL BLOMMERS
CHESTER W. HARRIS
HAROLD E. JONES

H. H. REMMERS
PERCIVAL M. SYMONDS

MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY



Baltimore, Md.
WARWICK and YORK, Inc.
1956

Bureau Ednl. Res. Research
 LAYTON HALL, ALABAMA COLLEGE
 Listed
 Acct. No.

Bureau Ednl. Res. Research
 DAVIS
 Listed 1-9-70
 Acct. No. J294

TABLE OF CONTENTS

American Doctoral Dissertations on Educational Psychology in Foreign Countries. WALTER CROSBY EELLS.....	133-136
Association Pathways in Language Disabilities. THOMAS H. EAMES.....	8-10
Behavioral Control and Intellectual Achievement of Secondary School Boys. NORMAN TALLENT.....	490-503
Book Reviews. 60-64, 125-128, 190-192, 251-256, 317-319, 380-383, 443-448, 504	
Boys Who Run Away from Home. WILLIAM W. WATTENBERG.....	335-343
Class and Sex Differences in Verbal Facility of Very Bright Children. HELEN H. DAVIDSON AND DOM BALDUCCI.....	476-480
College Teaching Ability, Scholarly Activity and Personality. A. H. MASLOW AND W. ZIMMERMAN.....	185-189
Comparison of the Factorial Structure of Cognitive Functions for a High and Low Status Group, A. JAMES V. MITCHELL, JR.....	397-414
Cross-Validation of a Work-Attitude Scale from the MMPI, A. RALPH TESSENEER AND MARY TYDLASKA	1-7
Davis-Eells Test and Reading Success in First Grade, The. IVAN L. RUSSELL.....	269-270
Determining the Relationship Between Individual Interest Profiles and Occupational Norms. ALBERT J. KINGSTON, CLAY E. GEORGE, AND W. PRICE EWENS.....	310-316
Development of New Scales for the Aptitude and Advanced Tests of the Graduate Record Examinations, The. MARGARET K. SCHULTZ AND WILLIAM H. ANGOFF.....	285-294
Differences Between Girl and Boy "Repeaters." WILLIAM W. WATTENBERG.....	137-146
Directed versus Independent Discovery of Established Relations. ROBERT C. CRAIG.....	223-234
Educational Theory and the Psychology of Learning. G. T. BUSWELL.....	175-184

Effect of Angular Alignment upon Readability of Print.	
MILES A. TINKER.....	358-363
Effect of Immediacy of Knowledge of Correctness of Response upon Learning.	
ARTHUR C. F. GILBERT.....	415-423
Effect of Teacher-Student Interviews on Classroom Achievement.	
ARTHUR J. HOEHN AND ELI SALTZ.....	424-435
Effects of Coöperation and Competition on the Cohesiveness of Small Face-to-Face Groups.	
N. PHILLIPS AND LOUIS A. D'AMICO.....	65-70
Effects of Non-promotion on Educational Achievement in the Elementary School.	
WILLIAM H. COFFIELD AND PAUL BLOMERS.....	235-250
Evaluation of a Student-Centered and Instructor-Centered Method of Conducting a Graduate Course in Education.	
AN. GLEN R. RASMUSSEN.....	449-461
Experimental Comparison of Nonsense, Word, and Sentence Materials in Early Typing Training.	
AN. LEONARD J. WEST.....	481-489
Fact and Fantasy in Vocational Interest Measurement.	
HILDEGARDE JOHNSON.....	40-46
Factor Models.	
A. E. MAXWELL.....	129-132
F-Scale, GAMIN, and Public School Principal Behavior.	
VYNCE A. HINES.....	321-328
General Ability to Judge Sociometric Status: Elementary Student Teachers' Sociometric Perceptions of Classmates and Pupils.	
THE. NORMAN E. GRONLUND.....	147-157
General Response Pattern to Five-Choice Items.	
EDWARD L. CLARK.....	110-117
Generality of Teachers' Sociometric Perceptions: Relative Judgment Accuracy on Several Sociometric Criteria.	
NORMAN E. GRONLUND.....	25-31
Group Functioning under Negative Conditions.	
LEONARD A. OSTLUND.....	32-39
Group Therapy and Remedial Reading.	
AARON LIPTON AND ARTHUR H. FEINER.....	330-334
Investigation of the Relationship Between Personality and Vocational Interest.	
AN. WILLIAM R. MELTON, JR.....	163-174
Learning of Subject Matter in Teacher-Centered and Group-Centered Classes.	
THE. GERARD V. HAIGH AND WARREN SCHMIDT.....	295-301

Longitudinal Comparison of Student Attitudes toward Minorities, A. LEWIS E. ALBRIGHT, ARTHUR D. KIRSCH, C. H. LAWSHE AND H. H. REMMERS.....	372-379
Longitudinal Study of Sibling Resemblances in Intelligence and Achievement, A. SARAH M. SCHOONOVER	436-442
Minnesota Teacher Attitude Inventory in Evaluating the Teaching of Educational Psychology, The. MORRIS E. ESON.....	271-275
Nondeliberative Factors in Teaching. J. M. STEPHENS..	11-24
Note on Factor Models, A. CHESTER W. HARRIS.....	265-268
Note on the Use of a "Don't Know" Alternative in Multiple Choice Tests, A. MORTON P. FRIEDMAN AND EDWIN A. FLEISHMAN.....	344-349
Objective Measurement of a Teacher's Classroom Interactions, An. JOHN WITHALL.....	203-212
Personality Factors of Over- and Under-Achievers in Engineering. ELVA BURGESS.....	89-99
Problem Behavior in School. MERRILL T. EATON, LOUIS A. D'AMICO AND BEEMAN N. PHILLIPS.....	350-357
Publications Received.....	320
Reading-Center Approach within the Classroom, A. JOSEPH JACKSON.....	213-222
Reading Improvement as a Function of Student Personality and Teaching Method. DONALD E. P. SMITH, ROGER L. WOOD, JAMES W. DOWNER, AND ALTON L. RAYGOR.....	47-59
Recency and Type of Undergraduate Training and Decline of G.R.E. Performance with Age. R. TRAVIS OSBORNE AND WILMA B. SANDERS.....	276-284
Relation of Intelligence and Physical Maturity to Social Status of Children. RICK F. HEBER.....	158-162
Relationship of Success in Lower and Higher Level Military Schools. JAMES H. L. ROACH.....	329-334
Research Checklist in Educational Psychology, A. PERCIVAL M. SYMONDS.....	100-109
Sex Differences in the Distribution of Teacher Approval and Disapproval Among Sixth-Grade Children. WILLIAM J. MEYER AND GEORGE G. THOMPSON.....	385-396
Some Effects of Short Term Training in Reading under	

Conditions of Controlled Motivation. JAMES C. REED.....	257-264
Some Relationships Between Methods of Instruction, Personality Variables, and Problem-Solving Behavior. IRVING MALTZMAN, EUGENE EISMAN, AND LLOYD O. BROOKS.....	71-78
Stimulus Familiarization as a Factor in Ideational Learning. WALTER WEISS AND BERNARD J. FINE.....	118-124
Student Achievement as a Measure of Instructor Effectiveness. JOSEPH E. MORSH, GEOEGE G. BURGESS, AND PAUL N. SMITH.....	79-88
Study of the Family Background of the Gifted, A. WALTER B. BARBE.....	302-309
Study of Student Teachers' Attitudes Toward Children, A. MERRITT C. OELKE.....	193-198
Study of Student Withdrawal at "W.U," A. EMMA W. BRAGG.....	199-202
Teacher-Pupil Rapport and Student Teacher Characteristics. GALEN W. DODGE AND DONALD O. CLIFTON.....	364-371
Theory Development and the Study of Teacher Behavior. DAVID G. RYANS.....	462-475

BOOK REVIEWS

Leopold Bellak. <i>The Thematic Apperception Test and the Children's Apperception Test in Clinical Use.</i> (A. S. EDWARDS).....	317
Arnold Bernstein, <i>On the Nature of Psychotherapy.</i> (A. S. EDWARDS).....	443
Albert H. Cohen. <i>Delinquent Boys: The Culture of the Gang.</i> (A. S. EDWARDS).....	382
Emery I. Gondor, <i>Art and Play Therapy.</i> (A. S. EDWARDS).....	443
Starke R. Hathaway and Elio D. Monachesi. <i>Analyzing and Predicting Juvenile Delinquency with the MMPI.</i> (H. MELTZER).....	63
Nelson B. Henry, Editor. <i>Mental Health in Modern Education.</i> (A. S. EDWARDS).....	444
Leonard Kornberg. <i>A Class for Disturbed Children: a Case Study and Its Meaning for Education.</i> (A. S. EDWARDS).....	318
Ruth Kotinsky and Helen L. Witmer, Editors. <i>Community Programs for Mental Health.</i> (A. S. EDWARDS)...	383
Paul V. Lemkau. <i>Mental Hygiene in Public Health.</i> 2nd Ed. (A. S. EDWARDS).....	60
V. J. McGill. <i>Emotions and Reason.</i> (H. MELTZER)....	504
Cecil V. Millard. <i>School and Child: a Case History.</i> (C. M. LOUTTIT).....	252
Truman M. Pierce, et al. <i>White and Negro Schools in the South.</i> (A. S. EDWARDS).....	380
Edward Podolsky, Editor. <i>Music Therapy.</i> (H. MELTZER).....	191
Helen D. Sargent. <i>The Insight Test (a Verbal Projective Test for Personality Study).</i> (HENRY H. MORGAN)	190
Ernst W. Swanson and John A. Griffen, Editors. <i>Public Education in the South Today and Tomorrow.</i> (A. S. EDWARDS).....	251
Herbert A. Thelin. <i>Dynamics of Groups at Work.</i> (MATTHEW B. MILES).....	125
Clyde B. Vedder. <i>The Juvenile Offender.</i> (H. MELTZER)	63
J. E. Wallis Wallin. <i>Education of Mentally Handicapped Children.</i> (A. S. EDWARDS).....	61
Robert I. Watson. <i>Psychology as a Profession.</i> (A. S. EDWARDS).....	443
Paul Witty. <i>How to Become a Better Reader.</i> (MILES A. TINKER).....	127

INDEX TO AUTHORS

ALBRIGHT, LEWIS E.....	372	LAWSHE, C. H.....	372
ANGOFF, WILLIAM H.....	285	LIPTON, AARON.....	330
BALDUCCI, DOM.....	476	MALTZMAN, IRVING.....	71
BARBE, WALTER B.....	302	MASLOW, A. H.....	185
BLOMMERS, PAUL.....	235	MAXWELL, A. E.....	129
BRAGG, EMMA W.....	199	MELTON, WILLIAM R., JR.....	163
BROOKS, LLOYD O.....	71	MEYER, WILLIAM J.....	385
BURGESS, GEORGE G.....	79	MITCHELL, JAMES V., JR.....	397
BURGESS, ELVA.....	89	MORSH, JOSEPH E.....	79
BUSWELL, G. T.....	175	OELKE, MERRITT C.....	193
CLARK, EDWARD L.....	110	OSBORNE, R. TRAVIS.....	276
CLIFTON, DONALD O.....	364	OSTLUND, LEONARD A.....	32
COFFIELD, WILLIAM H.....	235	PHILLIPS, BEEMAN N.....	65, 350
CRAIG, ROBERT C.....	223	RASMUSSEN, GLEN R.....	449
D'AMICO, LOUIS A.....	65, 350	RAYGOR, ALTON L.....	47
DAVIDSON, HELEN H.....	476	REED, JAMES C.....	257
DODGE, GALEN W.....	364	REMMERS, H. H.....	372
DOWNER, JAMES W.....	47	ROACH, JAMES H. L.....	329
EAMES, THOMAS H.....	8	RUSSELL, IVAN L.....	269
EATON, MERRILL T.....	350	RYANS, DAVID G.....	462
EELLS, WALTER CROSBY.....	133	SALTZ, ELI.....	424
EISMAN, EUGENE.....	71	SANDERS, WILMA B.....	276
ESON, MORRIS E.....	271	SCHMIDT, WARREN.....	295
EWENS, W. PRICE.....	310	SCHOONOVER, SARAH M.....	436
FEINER, ARTHUR H.....	330	SCHULTZ, MARGARET K.....	285
FINE, BERNARD J.....	118	SMITH, DONALD E. P.....	47
FLEISHMAN, EDWIN A.....	344	SMITH, PAUL N.....	79
FRIEDMAN, MORTON P.....	344	STEPHENS, J. M.....	11
GEORGE, CLAY E.....	310	SYMONDS, PERCIVAL M.....	100
GILBERT, ARTHUR C. F.....	415	TALLENT, NORMAN.....	490
GRONLUND, NORMAN E.....	25, 147	TESSENEER, RALPH.....	1
HAIGH, GERARD V.....	295	THOMPSON, GEORGE G.....	385
HARRIS, CHESTER W.....	265	TINKER, MILES A.....	358
HEBER, RICK F.....	158	TYDLASKA, MARY.....	1
HINES, VYNCE A.....	321	WATTENBERG, WILLIAM W.....	137, 335
HOEHN, ARTHUR J.....	424	WEISS, WALTER.....	118
JACKSON, JOSEPH.....	213	WEST, LEONARD J.....	481
JOHNSON, HILDEGARDE.....	40	WITHALL, JOHN.....	203
KINGSTON, ALBERT J.....	310	WOOD, ROGER L.....	47
KIRSCH, ARTHUR D.....	372	ZIMMERMAN, W.....	185

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

January, 1956

Number 1

A CROSS-VALIDATION OF A WORK ATTITUDE SCALE FROM THE MMPI

RALPH TESSENEER and MARY TYDLASKA

McNeese State College

In a previous article (18) one of the writers described the item analysis and selection of a scale from the items of the MMPI (12) which were found to discriminate between a group of 60 air force service personnel classified as "poor work attitude" and 50 industrial employees classified as "good work attitude."

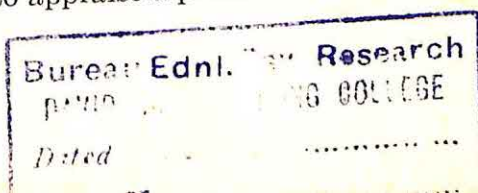
An item analysis provided items which significantly discriminated between the two groups, providing a Work Attitude scale. The writers believed that additional research would be desirable to validate these items as the possibility exists that work attitude may be found to be a situational factor rather than an independent syndrome.

PURPOSE OF STUDY

Recently, the writers had the opportunity to re-apply the Work Attitude scale to a different type of work attitude group and additional data were secured from a freshman and sophomore college population. A study was designed to evaluate the usefulness of the Work Attitude scale in determining motivation and underlying work attitude toward college achievement.

College faculty members commonly recognize that many students fall short of realizing their full capacities because of a lack of personal adjustment. The academic world is aware of the importance of underlying motivational and attitudinal factors in predicting scholastic achievement. Attempts to identify or to relate these motivational factors with academic success have not, in general, been particularly successful up to the present time.

In an effort to determine the validity of the Work Attitude scale in a homogeneous population and to appraise a presumed relation-



ship with scholastic achievement the following procedures were used to select subjects for study.

SELECTION OF SUBJECTS

Fifty-three faculty members at McNeese State College in Lake Charles, Louisiana, were requested by the writers to identify male white freshmen and sophomore students, with whom they were personally familiar, as outstanding examples of "good work attitude" and "poor work attitude." A total number of twenty-six students was identified as "poor work attitude" students by three or more faculty members. These subjects were selected for an experimental group. An equal number of subjects defined as "good work attitude" students by three or more faculty members was established. Criteria for equating were as follows: same age range, approximately the same proportion in class rank in college and equivalent mean scores on a general intelligence test. The "poor work attitude" group was composed of 16 freshmen and 10 sophomores, with a mean age of 21.3 years and whose mean I.Q. on the Otis Mental Ability Test A-m was 106.9. The "good work attitude" group was composed of 18 freshmen and 8 sophomores with a mean age of 20.7 and a mean I.Q. of 110 on the Otis Mental Ability Test A-m. A "t" test of significance of the means revealed that the groups did not differ significantly.

PROCEDURE

The total number of subjects, twenty-six in the "good work attitude" group and twenty-six in the "poor work attitude" group, were administered the group form of the MMPI and all tests were scored for an additional number of experimental MMPI scales as well as the Work Attitude scale. The following experimental scales were used: Do—Dominance (6), Sr—Social Responsibility (7), Sp—Social Participation (8), De—Dependency (15), Pr—Prejudice (5), Ht—Hyper Tension (11), Nm—Neuroticism (19), Lb—Low Back Pain (10) and Wa—Work Attitude (18). Table I presents the means and standard deviations for both groups of students.

An examination of the sub-scales in terms of their relationship to the two groups of individuals included an inspectional analysis of the MMPI profiles. This inspection was conducted in order to determine the number classified as normal and the number having

TABLE I—MEANS AND STANDARD DEVIATIONS OF THE MMPI SCORES FOR SAMPLES OF GOOD AND POOR WORK ATTITUDE COLLEGE STUDENTS MATCHED FOR INTELLIGENCE*

Scale	Good Work Attitude Group N = 26		Poor Work Attitude Group N = 26		't' Values
	Mean	Sigma	Mean	Sigma	
?	0	—	0	—	—
L	2.9	1.5	3.2	1.5	.69
F	2.9	2.1	4.4	3.9	1.72
K	14.8	4.1	13	3.8	1.55
Hs	50.6	6.7	51.5	10.5	.36
D	48.2	11.8	54.1	17.7	1.39
Hy	55.1	7.3	56.8	8.1	.77
Pd	54.9	8.9	61.2	8.7	2.31†
Mf	55.3	7.4	56.9	11.2	.61
Pa	51.8	6.6	56.5	8.6	2.18†
Pt	55.3	8.9	60.6	15.6	1.48
Sc	52.9	8.9	60.8	16.2	2.13†
Ma	54.6	7.5	59.6	12.1	1.72
Si	47.2	6.8	51.1	8.4	1.85
Do	16.2	2.6	13.8	3.6	2.73†
Sr	21.2	3.7	18.8	4.6	1.55
Sp	18.2	3.1	17.9	3.4	.32
De	18.4	5.9	23.2	9.5	2.18†
Pr	9.1	4.2	12.2	5.4	2.28†
Ht	13.6	4.3	14.1	7.4	.29
Nm	3.7	3.1	5.5	3.5	1.88
Lb	9.2	2.4	9.9	2.4	1.04
Wa	7.3	3.2	13.1	6.9	3.82†

* Intelligence determined by standard intelligence tests (Otis Gamma: Form Am). No significant difference between mean I.Q. scores of good and poor work attitude students.

† Significant at the .01 level of confidence.

‡ Significant at the .05 level of confidence.

T scores at 70 on one or more sub-scales. Table II presents the results of this inspectional analysis. Grade point averages for all students in both groups were computed and included a minimum of two semesters' work in college. Sufficient time elapsed between MMPI testing and computation of grade point average to permit grade point averages to be obtained for students who were classified as freshmen at the time of testing.

TABLE II—INSPECTIONAL ANALYSIS OF MMPI PROFILES OF
 "GOOD WORK ATTITUDE" STUDENTS AND "POOR WORK
 ATTITUDE" STUDENTS

	"Good Work Attitude"				"Poor Work Attitude"			
	N	%	Cumulative		N	%	Cumulative	
			N	%			N	%
Normal Profile	23	88	23	88	13	50	13	50
1 Sub-scale 70 or over	2	8	25	96	5	19	18	69
2 Sub-scales 70 or over					2	8	20	77
3 Sub-scales 70 or over	1	4	26	100	1	4	21	81
4 Sub-scales 70 or over					3	11	24	92
5 Sub-scales 70 or over					2	8	26	100
	26	100			26	100		

RELATED RESEARCH

A number of studies have reported norms for the MMPI for various college groups. Sopchak (17) has summarized the results of a number of investigators and also presented normative information for 316 male college students. He states "the group as a whole was normal according to the criteria of the authors of the test . . . however, at least 5 per cent of the men exceeded T scores of 70 on all scales except Hs, Hy and Pa." Atlas (1) item-analyzed the MMPI, searching for items predictive of academic achievement but his criterion groups were composed of both men and women so that item survival could be related to sex differentiation as well as scholastic achievement. A study of achieving and non-achieving college students by Morgan (14) found that achievers scored significantly higher than non-achievers on the MMPI experimental scales of Dominance, Social Responsibility and Intellectual Efficiency. On each of the scales the mean scores of both groups, while differing significantly from each other, were above the mean of the original norms for these scales.

RESULTS AND DISCUSSION

The mean scores for both groups of individuals on the MMPI scales with their respective standard deviations have already been given in Table I. Here it is noted that the representative group of twenty-six "poor work attitude" individuals displayed significant

differences on two of the MMPI experimental scales at the .01 level of confidence. Raw scores on the Work Attitude scale and on the Dominance scale for the "poor work attitude" group suggest that these scales are useful in determining motivation and underlying work attitude toward college achievement.

At the .05 level of confidence, significant differences were observed between the "poor" and "good" work attitude groups on the Pd, Pa, and Sc scales. The MMPI experimental scales, Dependence and Prejudice, differentiated at the .05 level.

The "poor work attitude" group shows a distinct, characteristic pattern which, in spite of the small numbers, is statistically significant. With respect to scores on the Work Attitude scale, the "poor" and "good" groups fall on opposite sides of the mean of a sample of the total college population. A similar disbursement of scores for both groups is noted on the Dominance scale. The Do scale implies such characteristics as "optimism, persuasiveness, self-discipline and resoluteness." (3)

These results if confirmed on another sample, indicate that the Work Attitude scale and the Dominance scale contribute to diagnosis. As far as an exploratory study permits we can roughly indicate the diagnostically useful pattern as follows: scores below the mean on the Dominance scale and scores in the critical range (11 and upward) on the Work Attitude scale with expressed tendencies on one or more of the Pd, Pa and Sc scales approaching significance.

This technique may provide counselors with more meaningful information from MMPI profiles within the normal range. Because of the amount of time and cost involved in counseling, any device which improves the efficiency of counseling services is highly desirable. The signs of latent personality disturbances, and in many cases, even existing maladjustment are often not revealed in overt behavior. A technique to aid counselors in identifying students whose work attitude may have a detrimental influence upon their scholastic achievement would be very useful.

Significant differences were found between the profiles of the "poor work attitude" students and the "good work attitude" students. For example, 13 (or 50 per cent) of the 26 "poor work attitude" cases had one or more T scores of 70 while only 3 (or 12 per cent) of the 26 "good work attitude" students had one or more scores of 70 or more.

The mean grade point average for the "good work attitude" group was 1.8 while the mean grade point average for the "poor work attitude" group was 1.1. This represents a difference of almost a grade interval between the two groups. A "t" test of significance revealed that the groups differed significantly at the .01 level of confidence ($t = 4.29$).

Work attitude may be a descriptive term for a general syndrome of behavior symptomatic underlying maladjustment. Subsequent studies would be required to specify the relationship, if any, between Work Attitude score and amount of anxiety present, level of aspiration or expressed valuations of self. It is possible that material from a projective instrument such as the TAT would be revealing.

SUMMARY

Two groups, each composed of twenty-six college students, who had been identified by three or more faculty members as examples of "good" and "poor" work attitude were administered the MMPI. Statistically significant differences at a high level of confidence were found between the groups on two experimental scales—the Dominance and Work Attitude scales. A tentative pattern to aid in the identification of "poor work attitude" risks (potentials) was established.

REFERENCES

- 1) W. D. Atlas, "A College Achiever and Non-Achiever Scale for the MMPI." *J. Applied Psychology*, 1948, 32: 385-397.
- 2) H. Brown, "Similarities and Differences in College Populations on the MMPI." *J. Applied Psychology*, 1948, 32: 541-549.
- 3) W. R. Dabson and D. R. Stone, "College Freshman Responses on the MMPI." *J. Ed. Research*, 1951, 44: 611-618.
- 4) H. G. Gough, "Factors Relating to the Academic Achievement of High School Students." *J. Ed. Psychology*, 1949, 40: 65-68.
- 5) H. G. Gough, "Studies of Social Intolerance: I, II, III, IV." *J. Social Psychology*, 1951, 237-246.
- 6) H. G. Gough, H. McClosky, and P. E. Meehl, "A Personality Scale for Dominance." *J. Abnormal & Social Psychology*, 1951, 46: 360-366.
- 7) H. G. Gough, H. McClosky, and P. E. Meehl, "A Personality Scale for Social Responsibility." *J. Abnormal & Social Psychology*, 1952, 47: 73-80.
- 8) H. G. Gough, "Predicting Social Participation." *J. Social Psychology*, 1952, 32.
- 9) J. P. Guilford, *Fundamental Statistics in Psychology & Education*. McGraw-Hill Book Co., New York, 1950.

- 10) L. J. Hanvik, "MMPI Profiles in Patients with Low-Back Pain." *J. Consulting Psychology*, 1951, 15: 350-353.
- 11) R. E. Harris, L. Carpenter, and M. Freedman, Unpublished study at the U. of California. Langlely Porter Psychiatric Clinic.
- 12) S. R. Hathaway and C. McKinley, *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press, 1943.
- 13) O. M. Lough, "Teachers, College Students, and the MMPI." *J. Applied Psychology*, 1946, 30: 241-247.
- 14) Henry Morgan, "A Psychometric Comparison of Achieving and Non-Achieving College Students of High Ability." *J. Consulting Psychology*, 1952, 16: 293-298.
- 15) L. Navran, "A Rationally Derived MMPI Scale for Dependence." Unpublished Ph.D. Thesis, Stanford Univ., 1951.
- 16) W. A. Owens and Wilma C. Johnson, "Some Measured Personality Traits of Collegiate Under-Achievers." *J. Ed. Psychology*, 1949, 40: 41-46.
- 17) Andrew L. Sopchak, "College Student Norms for the Minnesota Multiphasic Personality." *J. Consulting Psychology*, 1952, 16: 45-48.
- 18) Mary Tydlaska and Robert Mengel, "A Work Attitude Scale for the MMPI." *J. Applied Psychology*, 1953, 37: 474-477.
- 19) J. F. Winnie, "A Scale for Neuroticism: An Adaption of the MMPI." *J. Clinical Psychology*, 1951, 7: 117-122.

ASSOCIATION PATHWAYS IN LANGUAGE DISABILITIES

THOMAS H. EAMES

School of Education, Boston University

An area in the physiology of reading that has not been fully explored is the relationship of the tracts that make possible associations between the language centers in the brain.

It is well established in physiology that the language functions have a certain localization, with such factors as reading, hearing, speech, and even writing occurring around definite centers which have been demarcated and numbered by various investigators, notably by Brodmann (1).

Language function is controlled by and language memories are laid down in these areas. They are not isolated but have a high degree of interdependence. The remedial teacher is taking advantage of this arrangement when she makes use of as many sensory channels as possible in teaching reading—visual, auditory, speech and kinesthetic, for example. She is laying down simultaneous memories in the pupil's different language and associated areas such that any of a number of possible related stimuli will later awaken the desired memory of material learned. She is depending on association of visual, auditory, speech and kinesthetic learning to reinforce one another by contributing memories to form a whole.

The localized centers for language tend to be separated by silent or association areas where sense impressions are synthesized into complex perceptions. To and through these course tracts of fibres which connect with the various centers and probably with many others. An example of this is the "Language Formation Area" (posterior part of area 22 and upper part of 37), lying between the reading (39) and hearing (41-42) areas.

Representative of thinking in regard to the function of this area is the statement of Cobb (2) to the effect that a lesion here may cause as much reading difficulty as one in the reading area itself. This results from interruption of the association reflexes involving the pathways connecting the component language areas. Interference with conduction through any other tracts connecting lan-

guage and association areas may likewise impair language function. Therefore the integrity of the tracts are of as much practical importance to the student of language disabilities as the centers themselves.

While some of the tracts are known from histological studies and others can be seen macroscopically in anatomical sections, many are deduced from observed behavior and probably many are still unknown.

A few representative tracts are discussed with relation to language. For example, auditory memory loss may result from defects, trauma or disease of the long fronto-occipital fasciculus in the lower part of the external capsule. This bundle of fibres runs between the frontal, lower temporal and occipital lobes of the brain. Lesions of the dorsal superior longitudinal bundle of the superior frontal fasciculus may cause failure to associate auditory and visual memories correctly, since these fibres run between the inferior parietal lobe where the angular and supramarginal gyri (concerned with symbolization of words and digits) are located to the middle and inferior frontal convolutions which include the area associated with motor-speech.

Impairment of eye movements, including those made in reading, may result from lesions affecting, among others, the posterior longitudinal bundle of the median longitudinal fasciculus, the aberrant pyramidal fibres or the fibres of the superior cerebellar peduncle. Sometimes lesions of the superior occipito-frontalis are believed to prevent adequate associations between vision, eye movements and speech.

The visual fields, which bear a close relationship to eye span in reading, may be restricted in various ways by difficulties involving the optic radiations. Poor visual memory is likely to result from lesions of the short association fibres, connecting the various structures in the visual areas of the occipital lobe, particularly those connecting areas 17 and 18 with area 19, while reading difficulty can result when the fibres connecting these areas with area 39 are involved in anomalous processes.

The corpus callosum is a structure containing a large number of tracts which pass from one hemisphere to the other and probably reach every gyrus on both sides. While no positive evidence appears to exist to substantiate this, it seems possible that lesions affecting this structure might impede inter-hemispheric association

pathways enough to influence the condition of lateral dominance. This would be of considerable interest to students of reading, spelling, writing and speech difficulties.

While some lesions affecting the pathways, such as tumors, may be treated surgically, neither neuro-surgeons nor neuro-psychiatrists can do very much about improving conditions in the tracts at the present time. One of the purposes of this paper is to direct the attention of such persons to this promising field where more research is greatly needed. While complete understanding may have to wait for improvements in methods of histological study, it is likely that much can be accomplished with present methods and equipment.

For the present the practicing psychiatrist, clinical psychologist and teacher must continue remedial teaching, patiently, and with a greatly increased number of practice periods and motivational aids as compared to those required in teaching the child without language disability. Although much has been done by this group in terms of special methods adapted to the various language handicaps, there is still a great need for further research here as well as in the basic neuro-physiology of language.

REFERENCES

- 1) K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde*. Leipzig, J. A. Barth, 1909.
- 2) S. Cobb, *Borderlands of Psychiatry*. Cambridge, Harvard University Press, 1943, p. 43.

NONDELIBERATIVE FACTORS IN TEACHING

J. M. STEPHENS

The Johns Hopkins University

Teaching, as we have known of it for some centuries and in many cultures, is a highly deliberate process. It has been subjected to intensive analysis, and much of it is governed by thoughtful, deliberate decisions. The deliberate direction given to teaching is itself highly elaborate, and proceeds through an intricate arrangement of committees, boards and administrators.

It would be amazing if this elaborate, deliberate procedure did not have a considerable influence on the actual output of the schools. Undoubtedly, much of what the schools accomplish is linked up, in some way, with the deliberate decisions made in the legislatures, in the offices of the school board, the committee rooms or in the classroom. Without trying to deny the importance of these deliberate measures, however, it is still possible to suggest that, underlying these deliberate and elaborate procedures, there may be a few primitive, spontaneous forces which also contribute to educational growth. We have in mind the kind of forces that would be present even in the absence of any deliberate intention to teach, and that would prevail as a crude, constant substratum of forces in spite of dramatic fluctuation in deliberate, rational decisions.

In other discussions (23, 24) it has been held that the very existence of schools or of schooling derives, partially at least, from a postulated group of blind, spontaneous tendencies that are found in almost all human societies. In this current paper we maintain that these same tendencies which are partly responsible for the existence of schools, also operate, through nature's admirable parsimony, to produce a moderate degree of effectiveness in the work that the schools actually accomplish.

THE ANALOGY OF THE TYPICAL HOME

The non-deliberate components which we claim to find in the classroom are to be seen even more clearly and dramatically in the case of the educational work of the home. Through the activities of the typical home, most children learn a new language and acquire a large number of concepts and ideas. It is held that, in most homes,

the activities responsible for these attainments are largely spontaneous and automatic. Few of these activities are set in motion by the deliberate attempts of parents to provide instructions in the vernacular, or in table manners, or in the art of dressing. A few parents, it is true, do take a highly deliberate attitude toward these matters, and such parents may go in for considerable systematic instruction, but typically only a small part of the daily activity in the home is concerned with deliberate teaching. Most of the daily contact between parents and children, on the contrary, is concerned with the sheer business of dealing with matters of the moment. We talk to children, not primarily to give them practice in listening to oral speech, but to get them to come to the table, to get them to leave the cat alone, or because we merely happen to feel like commenting on the weather and the children are convenient listeners. Now it happens that these things *do* give the child practice in listening to oral speech, but only a small proportion of our utterances are motivated by this thought. Similarly when we express our delirious joy at the child's pronouncement of "da da," we do so, not primarily because of a deliberate desire to encourage him to repeat the performance in the future, but simply because we feel joyful and give spontaneous expression to our ecstasy. Conversely, when we express consternation upon seeing him upset his soup, we do so, not necessarily to discourage this in the future, but chiefly because we feel worried or distressed and spontaneously show it. We show it even when we feel we ought not.

SPONTANEOUS TENDENCIES HELD TO BE AT WORK IN EACH CLASSROOM

The spontaneous tendencies which would tend to produce the institution of schooling and which are seen to some extent in the activities of the home are undoubtedly numerous and are probably highly specific. For convenience in description, however, these many individual tendencies have been grouped into a number of clusters. Each of these clusters is merely a construct, of course, and is not supposed to represent some single entity within the organism. The clusters of these postulated tendencies are as follows:

- 1) A cluster of tendencies responsible for our playful, manipulative behavior (such as fishing, stamp collecting, bird watching, solving cross-word puzzles, or collecting Lincoln anecdotes). The essential feature of these activities is that, ordinarily, they do not

have any immediate, clear-cut survival value for the individual or the race, but that they are interesting and that, in the long run, a very few of them may turn out to have important survival value. It is held that many of these playful tendencies are basic and widely prevalent and are to be taken seriously in their own right.

2) A cluster of spontaneous tendencies leading us to talk to other people about our interests. Here the emphasis is not on our deliberate efforts to help or improve the hearer, but on our spontaneous description of our trips, operations, pets, golf scores, and of the things that we happen to know.

3) A cluster of spontaneous tendencies leading us, on the one hand, to tolerate, approve, or applaud some performance of other people and leading us, on the other hand, to express our disapproval of the performances of others, or to correct an inappropriate performance, or to show discomfort upon its occurrence.

Here again we do not refer to the conscious, altruistic efforts to encourage other people, or to increase their competence, but to the spontaneous, automatic smile, or applause, or to the uncontrolled grimace, or to the spontaneous unthinking correction that springs to our lips when we hear a word mispronounced, or detect a factual error.

4) As a combination of 1 and 2, we suggest a cluster of spontaneous tendencies which combine to lead us to "point the moral," to call attention to the things that follow from a given course of action. ("See what happens when you forget your age." "See how much better things are when you follow Mother's advice.") It is held that a great many of these comments do not arise from conscious desire to reform our listeners. It is held, on the contrary, that many of them spring spontaneously to our lips often even in the face of a former resolve to remain silent.

These tendencies are held to be almost universally prevalent and are by no means confined to the school or the home. Indeed such tendencies may perhaps be seen more clearly at work in the bird-watcher, the golfer, or the amateur historian. Here we find people greatly interested in things that have no discernible immediate value of a practical nature. From among these people, moreover, we can find quite a few who are quite vocal about their interests and who will hold forth on their hobbies to any available listeners, often from sheer compulsion to talk, often with no thought of helping the listener, often in the face of the listener's obvious boredom.

In some of these enthusiasts, the spontaneous tendency to react to errors may also be present to a marked extent. A few ornithologists, for instance, may readily wince, squirm, or supply a correction when they hear a heron called a crane. The amateur sailor may react similarly when he hears the "deck" referred to as a "floor." Again in these spontaneous compulsions to offer a correction, there is no necessary concern for the welfare of the erring one. The correction or display of distress is often quite unthinking and indeed may go counter to a considered resolve.

Although these spontaneous tendencies are assumed to be present to some extent in everyone, it is also assumed that they are stronger in some people than in others. It is further held that, on the average, teachers are not lacking in these tendencies. On the average, teachers, more than most people, are interested in various subjects that have much appeal but only indirect, or long range, practical value. By accident, or by design, or by tradition, teachers are a selected group with respect to their interests in reading, in numerical relations, in history, and in science. For the most part, moreover, teachers have at least an average tendency to hold forth on the subjects of their interest. In addition to speaking of these subjects for pay or as a part of their duty, many teachers will hold forth with little urging, apparently sharing the compulsion of the dog owner to speak of his pet, or the mother to wax eloquent on the doings of her child. The compulsive tendency to correct the mistakes of others, if there is any truth whatever in the literary stereotype, is certainly not lacking in teachers. "Teacher" is the epithet that Eliza Doolittle hurls at Professor Higgins when she is driven beyond endurance by his repeated corrections. The stereotype may be wrong in claiming that we are unusually endowed with this tendency, but we are not under-equipped in this respect.

It is the contention of this paper that a person liberally endowed with the tendencies in these clusters, and spending much time in the presence of children would be bound to induce some educational growth in these children, even if he had no intention of inducing such growth. Let us, for instance, imagine a man who has an exceptional interest in some such subject as mathematical puzzles, bird lore, the life of Lincoln, or Greek mythology. Let us suppose also that he ranks somewhat above average both in his tendency

to talk about his special interest and in his spontaneous tendency to commend or correct the actions or comments of others with respect to his special field. Now let us imagine that such a man is required to "baby-sit" with growing children for an hour or so each day. But in all this, let us insist that this man has not been assigned to teach the children about the subject of his interest, and that he himself has not overtly assumed that intention.

THE ABILITY OF THE POSTULATED TENDENCIES TO ENGAGE THE BASIC MECHANISMS OF LEARNING

Even when the man himself has no intention to instruct the children, it is held that the operation of the spontaneous tendencies would be bound to engage the essential mechanisms of learning and to bring about some degree of educational growth. It is held, conversely, that the basic mechanisms of learning are remarkably responsive to the pressures bound to be applied by the postulated spontaneous tendencies. Such a "gearing together," of course, is not surprising, since, in the history of the race, the acquisition of the tendencies and of the mechanisms necessary for learning probably evolved hand in hand.

THE BASIC MECHANISMS OF LEARNING

In showing that the postulated tendencies are likely to engage the basic mechanisms of learning, we are considerably handicapped, of course, by the disagreement on the nature of those basic mechanisms of learning. In the face of this spirited disagreement, we shall not try to choose between the many mechanisms that have been proposed. We shall take a safer, if less courageous, course and try to show that our postulated spontaneous tendencies, if present in adults, would engage any of the learning mechanisms that are currently considered and would do so to some extent even in the absence of any intention to teach. The following mechanisms, obviously interrelated and far from discrete, have been stressed by one theorist or another in recent years: (2, 8, 16, 19, 25)

1) *Motivation or Drive*—Learning is more likely to take place when some strong need, drive or pressure is acting on the organism. For some theorists the chief function of motive is to induce the activity necessary for learning. For other theorists the presence of some need is also a necessary condition for reinforcement (10).

Still others (26) do not consider motivation necessary for learning but consider it necessary for the behavior that demonstrates the fact of learning.

2) *Practice or Experience*—This is not always stressed in contemporary theory, but even when not stressed it is almost always implied or taken for granted. In some theories (10), repeated experiences are important in their own right. In others these repetitions permit more intrinsic forces to operate.

3) *Reinforcement or Effect*—According to some theorists, learning is more likely to occur when different ways of behaving are followed by different conditions (need reduction or lack of need reduction, confirmation of an expectancy, or refutation of an expectancy, satisfaction or annoyance).

This is the central doctrine in some theories (10). According to other theorists, however, reinforcement does nothing in and of itself, but must work through some other mechanism (5, 26).

4) *Insight, Cognition Formation, or Restructuring the Experiential Field*—Learning is more likely to occur when the individual acquires an insight or "cognitive map"; when he comes to see more clearly that a certain way of behaving leads to a certain result. This is the basic doctrine of Tolman (26) and of some of the classical Gestaltists (11).

5) *Substitution, Guidance, or the Conditioned Response Paradigm*—The learner will be more likely to perform the response "*R*," in the presence of "*S*" if, when he originally encounters *S*, he is led or forced, perhaps by some other means (*S*₁), to perform *R*.

This principle is the central feature of Guthrie's theory (5) and it plays a varying part in the different formulations put out by Hull (10).

These five mechanisms are by no means completely independent. In invoking any one of them, we are almost bound to find that we have also called some of the others into play or, indeed, perhaps all of the others. Many theories, as a matter of fact, (13, 15, 20, 21, 27) are known as two-phase theories or as mediation theories because of their attempt to integrate two or more mechanisms.

Let us now consider whether or not each of these mechanisms is likely to be invoked by (say) an adult Lincoln enthusiast who is liberally endowed with the postulated tendencies but who does not intend to teach.

MOTIVATION

The motivation induced by our Lincoln fan will probably be crude and extrinsic when judged by some of the current standards (9). But it is difficult to think that the children will not be motivated, or led, or driven to listen and to react to the ideas propounded. Depending somewhat on the prestige of the adult, there may also be considerable need to gain his approval or good opinion. Most of us, when we hear a subject discussed frequently and earnestly by people whom we take seriously, feel somewhat inferior and unhappy if we know nothing of the subject, and less unhappy if we can follow with some comprehension. Under these circumstances the children may find themselves under pressure to be able to "go along" with the talk they hear.

PRACTICE

The children will inevitably have some experience or "practice" with regard to Lincoln. They will be bombarded by Lincoln statements. To many of those statements they will react silently. To others they make some overt comment. But Lincoln concepts are bound to be a part of their experience. Although this may be a far cry from the experience-curriculum, it is still experience of a kind and of a kind that will call for some reacting on their part.

REINFORCEMENT

The Lincoln enthusiast, although innocent of any intention to teach, is still likely to apply a good deal of reinforcement. Deliberate intention to help or to teach is not necessary to make sure that the mechanism of reinforcement will come into play. A true Lincoln zealot, plentifully endowed with the postulated spontaneous tendencies, would be bound to make some comment if he heard an erroneous statement about Lincoln. He would make such a comment whether or not he had any intention of helping the one who made the mistake.

If overt corrections of erroneous statements are often spontaneous rather than deliberate, then certainly the minor reactions to such statements are even more likely to be automatic or spontaneous. Without trying to, we look incredulous or bewildered or surprised when we encounter behavior that does not fit our standards. Conversely we beam with pleasure when we meet reactions

which are in line with our hopes and expectations. These subtle automatic expressions of approval or disapproval, like the overt expressions discussed above, do not need to be cultivated. On the contrary, they are extremely difficult to suppress. We must work hard to retain a poker face when someone is taking liberties with our specialty. Even then, the more discerning can detect slight expressions of dismay when the provocations become extreme.

These automatic or non-deliberative reinforcements, by the way, are by no means confined to the chance adult who lacks the intention to teach. They probably play an enormous rôle in the everyday work of the "intentional" teacher. For every time that any teacher marks a question right or wrong, or administers an overt expression of approval or disapproval, there are scores of occasions when his face shows a look of surprise, bafflement, incredulity, patient waiting, or relieved acceptance. These subtle grimaces, these shadows and lights, these nuances in tone of voice, are in play day in and day out as the teacher faces the class. They are very effective reinforcements. They come and go with very little deliberate forcing on the part of the teacher. And yet, it is quite possible that they contribute enormously to educational growth.

Even when the captive audience is listening "passively," as in the much condemned lecture situation, there are many automatic reinforcements. Actually, of course, mere listening is seldom completely passive. Most of us in listening (or reading), tend to anticipate the ideas that are about to come. The beginning of the sentence leads us to expect or, to guess, a certain conclusion. In this very pedestrian way, at least, we actively respond to the ideas that are presented. If the speaker or writer concludes the sentence or paragraph in accord with our guess or expectation, we experience one kind of reinforcement. If the conclusion is contrary to our guess or expectation, the reinforcement is quite different and may be moderately jarring or disconcerting.

ENHANCEMENT OF INSIGHT

Our insights or cognitive maps do not arise exclusively, or primarily, from the verbal comments of those around us. More often, presumably, they arise as the result of our own experiences. There seems no question, however, but that these experience-induced insights can be enhanced or sharpened by the comments of others. In exclaiming, "See, the cat doesn't growl when you stroke him

this way!", the mother may sharpen and clarify a "means-end Gestalt" that would otherwise be only partially available to the child. Similarly a host of other maxims, rules, or principles may enhance and "structure" cognitions or insights that otherwise would be only dimly apprehended.

In our postulate, of course, it is held that much of the tendency to point the moral, to sharpen the insight, is completely spontaneous and wholly devoid of altruistic intent. An adult associating with children would have to work hard not to call attention to the fact that this sort of action leads to that kind of result. We have to exercise continual restraint not to "rub it in" so to speak and to refrain from pointing out how much better things are now than the youngster has (finally) adopted this procedure rather than that.

SUBSTITUTION OR GUIDANCE

The postulated tendencies are likely to invoke the mechanisms of motivation, practice, reinforcement, and insight enhancement. Are they also likely to engage the mechanism of substitution, guidance, or conditioned-response learning?

Conditioned response learning occurs when the organism, in the presence of a given situation ($3 + 4 = ?$), is led, or guided, or forced to make a given response (saying or thinking 7). The rôle of guidance may be shown, in over-simplified form, by the following examples:

Little or No Guidance

S1	R1	S2	R2
Hears: $3 + 4 = ?$	Says: 12	Hears: No!	Says: (? ?)

Guidance Present

S1	R1	S2	R2
Hears: $3 + 4 = ?$	Says: 12	Hears: No! It's 7.	Says: Oh sure! 7.

The first example consists of pure trial-and-error learning. Reinforcement is provided, but guidance is almost completely lacking. In the second example, however, there is an element of guidance,—a very mechanical kind of guidance, it is true—but a kind that may be extremely effective. In the second example, the teacher, instead of merely saying "No!", goes on to say "It's 7," and thereby introduces something which leads the child to think of 7. The second example, which includes this extremely simple element of guidance, constitutes a typical illustration of conditioned-response learning.

According to the substitution or contiguity theories, S1 ($3 + 4 = ?$) will now become connected with R2 (saying 7).

It is held that the tendency to supply the correct answer, and thus invoke the mechanism of substitution, is quite spontaneous. Upon hearing a child give 12 as the answer to $3 + 4$, it is the most natural thing in the world not only to indicate that it is wrong but to go on and supply the correct answer. Consistent failure to do so, when observed, would suggest deliberate suppression of a spontaneous tendency. This unthinking, spontaneous urge to supply the correct answer when known is very wide-spread. This is the spontaneous urge that the director of the quiz program attempts to suppress when he pleads for no prompting. It is the urge often seen in the frantic hand waving of school children who seek to correct the mistake made by one of their colleagues. It is not necessarily a deliberate attempt to help or reform the erring one. The mistake is detected, the answer is known, and powerful, primitive urges automatically go into action.

All in all, it would seem that the postulated spontaneous tendencies, if they exist, would be bound to be called into play whenever an adult possessing them is supplied with a captive audience. The operation of these tendencies is bound to engage the mechanisms of learning here listed and thus to induce a measure of scholastic growth.

IMPLICATION

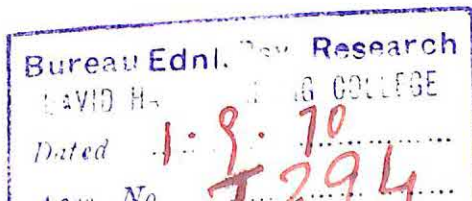
To extol the potency of these primitive non-deliberate forces is not to disparage the more deliberate, intelligent forces that may be added to them. Insofar as we are aware of these steady, hidden, earthy forces, however, we might take a different attitude toward the gains to be expected from the modification of the deliberate processes used by the schools. So long as we believe that scholastic growth comes entirely or almost entirely from these deliberate processes, it would be most natural to expect that an improvement in these processes would lead to a considerable gain in the scholastic growth of the children. Our expectations would surely change, however, if it appeared that a substantial amount of growth were largely independent of deliberate efforts. Under these circumstances we would probably set our expectations at a much more modest level.

In suggesting more modest expectations from educational in-

novations, the theory of spontaneous forces may have much in its favor. Actually scholastic growth has not seemed to be especially responsive to scholastic innovations. On the contrary, scholastic growth has exhibited a surprising constancy in the face of many drastic changes in policy or in deliberate procedures. Ever so often we adopt new philosophies or methodologies. At the very least, we seem to chorus new slogans. And yet, an observer might find it hard to detect any clear-cut changes in the more narrow scholastic growth attributable to changes in official policy. He might get the impression, on the contrary, that scholastic growth in the traditional subject matter continued at fairly steady pace and stubbornly refused to coöperate with the many decisions emanating from the conference room.

The problem of the constancy of scholastic growth is, of course, only an incidental item as far as the theory of spontaneous forces is concerned. Nevertheless some extreme claims for such constancy are made and some justification may be in order. At one time it was assumed as a matter of common knowledge that the preponderance of educational studies found little difference in the scholastic growth resulting from different procedures (17) and certainly this matter has been discussed, ever since Rice (18) discovered the surprising constancy of spelling attainment in the face of marked variations in the time devoted to study and since Merriam (14) reported the regular growth in school subjects in the absence of formal instruction in the subjects. These isolated reports tend to be supported by some of the more comprehensive summaries (7, 22).

At first glance, these assertions about the constancy of scholastic growth may seem to be in conflict with the many recent publications justifying the work of the modern school. For the most part, however, there is no actual conflict. The change so frequently levelled against the present-day school is not merely that modern education fails to produce superior results. The critics go much further and maintain that, as far as the "fundamental" scholastic subjects are concerned, the modern school is achieving much less than its predecessor. In view of this charge, it is enough for the champions of modern education to show that, again in the realm of traditional scholastic growth, the modern school is no worse than the older schools. And this, for the most part, is the tenor of the argument (3, 4, 6, 12). The champions maintain that the school



is *holding its own* in the matters of the three R's and that the achievement of present day pupils certainly equals that of earlier generations.

It would be wrong to imply that the defenders of modern education never claim an actual superiority. In many instances an apparent superiority is observed and the protagonists dutifully report it. The more cautious champions, however, are aware of occasional examples of evidence in the other direction, and typically they confine themselves to the very proper and important contention that modern-school, in spite of its concentration on a wider, real-life curriculum, has also been able to hold its own in the traditional field.

Obviously the recent defense of the school merely underlines and high-lights the relative constancy of scholastic results. Now we realize more clearly than ever that we have had important changes in the organization of the school which some people would expect to expedite mastery of the basic skills, and which others would expect to spell disaster for the basic skills. And yet, in spite of all these extensive changes, we find that development proceeds at about the same rate. When significant differences are announced they are still very slight, and typically any occasional advantage that may appear is promptly offset by some other chance departure in the opposite direction.

To one group of students this constancy of scholastic growth has seemed to be largely an illusion or artifact. It is suggested that this lack of change in results merely reflects the insensitivity of our tests, or of our experimental procedures. There is also some suggestion that this constancy comes from our traditional concentration on narrow scholastic growth. These latter suggestions imply, of course, that the broader aspects of growth might well prove more responsive to deliberate innovations or to changes in method.

Other students of education have assumed that there may be some truth behind the apparent results and have advanced hypotheses to account for the relative constancy. It has been suggested, for instance, that a considerable portion of educational growth is due to maturation (1), and that the growth due to this factor would necessarily fail to respond to the fluctuations in administrative policy or to other deliberate pressures. Similarly, it is suggested here that the non-deliberate components inevitably

at work in almost any teaching situation, will also help explain some of this apparent constancy.

SUMMARY

It is held that a group of postulated spontaneous tendencies, if they exist, would not only account for the establishment of institutions such as schools, but would also, in large measure, provide the actual tools by which the basic mechanisms of learning are engaged. These blind, automatic tendencies, which account for much of the learning that takes place in the home, are also inevitably at work when the more sophisticated teacher deliberately sets out to instruct. Underlying his deliberate intention, these automatic, spontaneous tendencies inevitably bring about motivation, practice, reinforcement, guidance, and the enhancement of insight. This statement, of course, is no real disparagement of the deliberate efforts that are added to the primitive forces. In the same way, it is no real disparagement of the physician to emphasize the many automatic mechanisms that are at work, before he applies his deliberate efforts, which may also contribute to recovery along with his more elaborate therapy. In both his case and in ours, such forces, if they exist, should not be fearfully disowned, but should be sought out, investigated and, if justified by further study, used as the basic pattern of forces on which our deliberate efforts might be imposed.

REFERENCES

- 1) S. A. Courtis, "The rate of growth makes a difference." *Phi Delta Kappan*, 1949, 30: 316-323.
- 2) J. E. Deese, *The psychology of learning*. New York, McGraw-Hill, 1952.
- 3) J. R. Gerberich, "The first of the three R's." *Phi Delta Kappan*, 1952, 33: 345-349.
- 4) W. S. Gray, "What is the evidence concerning reading?" *Progressive Educ.*, 1952, 29: 105-110.
- 5) E. R. Guthrie, *The psychology of learning*. New York, Harpers, 1952.
- 6) L. W. Harding, "How well are schools now teaching skills." *Progressive Educ.*, 1951, 29: 7-14, 32.
- 7) D. Harris, "Factors affecting college grades, a review of the literature." *Psychol. Bull.*, 1940, 37: 125-166.
- 8) E. R. Hilgard, *Theories of learning*. Appleton-Century-Crofts, 1948.
- 9) E. R. Hilgard and D. H. Russel, "Motivation in school learning." Chapt. 2 in *Nat. Soc. Stud. Educ. 49th yearbook*, Pt. 1, 1950, 36-68.
- 10) C. L. Hull, *Principles of behavior*. New York, Appleton-Century, 1943.
- 11) K. Koffka, *Principles of Gestalt psychology*. Harcourt Brace, 1935.

12) J. P. Leonard and A. C. Eurich, *An evaluation of modern education*. New York, Appleton-Century-Crofts, 1942.

13) N. R. F. Maier and T. C. Schnierla, "Mechanisms in conditioning." *Psychol. Rev.*, 1942, 49: 117-134.

14) J. L. Merriam, "How well may pupils be prepared for high school work without studying arithmetic, grammar, etc., in the grades?" *J. educ. Psychol.*, 1915, 6: 361-364.

15) O. H. Mowrer, "On the dual nature of learning—A reinterpretation of 'conditioning' and 'problem-solving'." *Harvard Educ. Rev.*, 1947, 17: 102-148.

16) C. E. Osgood, *Method and theory in experimental psychology*. Oxford Univ. Press, 1953.

17) C. C. Peters and W. R. Voorhis, *Statistical procedures and their mathematical bases*. New York, McGraw-Hill, 1940, p. 476.

18) J. M. Rice, "The futility of the spelling grind." *The Forum*, 1897, 23: 163-172.

19) K. W. Spence, "Theoretical interpretations of learning." Chap. 18 in S. S. Stevens' *Handbook of experimental psychology*. Wily, 1951.

20) B. F. Skinner, *The behavior of organisms*. Appleton-Century, 1938.

21) J. M. Stephens, "Expectancy vs. effect-substitution as a general principle of reinforcement." *Psychol. Rev.*, 1942, 49: 102-116.

22) J. M. Stephens, *The influence of the school upon the individual*. Edwards Bros., 1933.

23) J. M. Stephens, "Non-deliberative factors underlying the phenomenon of schooling." *Educational Theory* (In press).

24) J. M. Stephens, "The residual theory again: An analytical study." *Educational Theory* 1955, 5: 158-166.

25) L. P. Thorpe and A. M. Schmuller, *Contemporary theories of learning*. Ronald Press, 1954.

26) E. C. Tolman, *Purposive behavior in animals and men*. Century, 1932.

27) R. S. Woodworth, "Reinforcement of perception." *Amer. J. Psychol.*, 1947, 60: 119-124.

GENERALITY OF TEACHERS' SOCIOMETRIC PERCEPTIONS: RELATIVE JUDGMENT ACCURACY ON SEVERAL SOCIO-METRIC CRITERIA

NORMAN E. GRONLUND

College of Education, University of Illinois

There have been many investigations of social perception. Typically, they have been concerned with the accuracy with which persons judge the opinions (6), interests (3), personality traits (2), or other social characteristics of individuals or groups. Sociometric perception is considered a type of social perception in which individuals make judgments of the sociometric status of the members in a given group. Accuracy of perception is determined by comparing the judgments of sociometric status with the results of a sociometric test. Previous studies (4, 5) have shown that individuals vary greatly in the accuracy of their sociometric perceptions. Investigations have been made of the personality correlates of perceptual accuracy as well as the relation between accuracy of social perception and effectiveness in interpersonal relationships. These are both important areas of study, but the research results will be of limited value unless it can be established that perceptual skills in the social areas have some generality. If accuracy of social perception varies from one type of judgment to another, from one group to another, and from one situation to another, there is little basis for predicting the general accuracy of an individual's judgments of other people. If generality of social-perception skills can be established, their significance in interpersonal relationships can be more carefully studied and their correlates and trainability determined. Eventually it may be possible to select and train people in regard to this ability. This would be a worthy contribution in those areas of endeavor where accurate social perceptions would enhance successful performance. It would seem to be an especially important adjunct to the selection and training of teachers.

A few studies have suggested the possibility that there may be a general ability to judge others' behavior. Ausubel, Schiff and Gasser (1) analyzed the ability of public-school students to judge

the sociometric status of their classmates. Split-half reliability coefficients for judgments of others' sociometric status were .89 and .71 in the eleventh and twelfth grades, respectively. This indicates a fairly high degree of generality of such judgments over individuals. In a related area of social perception, Gage (3) had college students make judgments of others' interests. He obtained reliability coefficients of about .70 for generality over both individuals and test items. Travers (7) investigated the general ability of individuals to judge group knowledge. His findings revealed that the ability of college students to judge accurately the knowledge of their classmates was associated with the ability to judge accurately the knowledge of adults in general. Although these studies provide valuable leads concerning the existence of a general ability to judge others, further research is needed on all phases of the problem.

This study is an initial step toward the determination of the extent to which sociometric perception is a general ability. Specifically it is concerned with the generality of teachers' judgment accuracy of sociometric status over the criteria of seating companion, play companion, and work companion in forty sixth-grade classes. Generality, as used here, refers to the extent to which a teacher's judgment accuracy on one criterion is related to his judgment accuracy on each of the other criteria. Another study, in progress, is concerned with the generality of such judgments over different classroom groups. If a general ability to perceive sociometric status is established by these and other studies, future research can be directed toward a determination of the relation between sociometric perception and other forms of social perception.

METHOD OF INVESTIGATION

The data for this study were collected in connection with a previous investigation (4). The pupils in forty sixth-grade classes responded to a sociometric test requesting them to choose the five classmates they most preferred as seating companions, play companions and work companions. The sociometric status of the pupils was determined on each criterion by totaling the number of choices received. In the same forty classes each teacher made rank-order judgments concerning the relative sociometric status of the pupils in her class. These judgments were made separately for boys and girls on each of the three criteria.

TABLE I—ACCURACY OF TEACHERS' SOCIOMETRIC PERCEPTIONS: MEANS AND STANDARD DEVIATIONS OF 240 CORRELATION COEFFICIENTS REPRESENTING THE ACCURACY OF TEACHERS' JUDGMENTS OF THE SOCIOMETRIC STATUS OF BOYS AND GIRLS IN FORTY SIXTH-GRADE CLASSES

Criteria	Boys			Girls		
	N	M	SD	N	M	SD
Seating	40	.61*	.15	40	.62*	.17
Play	40	.57*	.20	40	.53*	.20
Work	40	.59*	.22	40	.64*	.17

* Significant at the 1 per cent level.

There were 632 boys and 626 girls in the forty sixth-grade classes. Class size ranged from fifteen to forty-three pupils with an average of thirty-two. All of the teachers were women.

ANALYSIS OF DATA AND RESULTS

Accuracy of sociometric perception was determined by correlating the teachers' rank-order judgments of the pupils' sociometric status with the results of the sociometric choosing. Since this was done separately for boys and girls on each of the three criteria, there were six correlation coefficients representing the accuracy of each teacher's judgments. The means of the 240 correlation coefficients resulting from this analysis are presented in Table I. An inspection of these means will reveal that the average accuracy of the teachers' sociometric perceptions is similar for boys and girls on all three criteria, with lowest accuracy on the play-companion criterion. The standard deviations in Table I indicate considerable variation among teachers in the accuracy of their judgments. Part of this variation can be accounted for by the spread of sociometric-status scores in the various classes. The standard deviations of the boys' sociometric-status scores in the forty classes ranged from 2.1 to 4.9 with a mean of 3.1. When correlated with accuracy of the teachers' judgments, a correlation coefficient of .27 was obtained. The standard deviations of the girls' sociometric-status scores ranged from 1.0 to 5.0 with a mean of 3.2. The correlation with accuracy of the teachers' judgments yielded a correlation coefficient of .34. The latter correlation coefficient was significant at the five per cent level. The lack of significance of the correlation between

the dispersion of boys' sociometric-status scores and accuracy of the teachers' judgments can probably be accounted for by the restricted range of standard deviations. These positive correlation coefficients indicate a tendency for the teachers' judgments to be more accurate where there is a greater spread of sociometric-status scores. However, the size of these correlation coefficients indicates that the variability of sociometric-status scores can account for only part of the variation in the accuracy of the teachers' judgments. The remainder of the variation can probably be attributed to the relative ability of the teachers to make such judgments.

Generality of Sociometric Perception—The accuracy scores of the teachers' judgments of sociometric status on the criteria of seating companion, play companion, and work companion were correlated, separately for boys and girls, to determine the generality of the accuracy of sociometric perception over criteria. The results, presented in Table II, reveal that all of the correlation coefficients are statistically significant at, or beyond, the five per cent level. The relatively small correlation coefficients representing generality of judgment accuracy over the play-companion criterion and the other two criteria can be accounted for, in part, by the variation among teachers in their opportunity to observe the pupils in play situations. Some of the teachers reported that they supervised their pupils during recess, while others remained in their classrooms and left the supervision to regular playground supervisors. The latter had little, if any, opportunity to observe choice behavior in play situations. The influence of this variable on the results cannot be

TABLE II—GENERALITY OF THE ACCURACY OF TEACHERS' SOCIOMETRIC PERCEPTIONS OVER CRITERIA: CORRELATION COEFFICIENTS REPRESENTING GENERALITY OF THE TEACHERS' JUDGMENT ACCURACY OVER THE CRITERIA OF SEATING COMPANION, PLAY COMPANION, AND WORK COMPANION IN FORTY SIXTH-GRADE CLASSES

Criteria	Boys		Girls	
	N	r	N	r
Seating-Play	40	.32†	40	.59*
Play-Work	40	.34†	40	.37†
Seating-Work	40	.52*	40	.67*

* Significant at the 1 per cent level.

† Significant at the 5 per cent level.

determined from the data, but in all probability it lowered the generality of judgment accuracy where the play-companion criterion was involved. Correlation coefficients representing generality of judgment accuracy over the criteria of seating companion and work companion were significant at the one per cent level for both boys and girls. These data are consistent with the hypothesis that the accuracy of teachers' sociometric perceptions has generality over criteria.

In an attempt to identify factors which might have influenced the amount of generality of the accuracy of teachers' sociometric perceptions, it was decided to compare the generality of the pupils' sociometric-status scores with the generality of the teachers' rank-order judgments, over the several criteria. This would indicate the extent to which the teachers erred in considering sociometric status a generalized trait or one that is specific to each criterion. The analysis of data, presented in Table III, indicates that the teachers tended to under-generalize in their judgments of sociometric status. The mean correlation coefficients representing generality of their judgments over criteria were, in all cases, smaller than the mean correlation coefficients representing generality of pupils' sociometric status scores. These differences between means were found to be statistically significant at the levels indicated in the table.

An inspection of the raw data has indicated that thirty of the

TABLE III—DIFFERENCES BETWEEN MEAN GENERALITY OF PUPILS' SOCIOMETRIC-STATUS SCORES AND MEAN GENERALITY OF TEACHERS' JUDGMENTS OVER THE CRITERIA OF SEATING COMPANION, PLAY COMPANION, AND WORK COMPANION IN FORTY SIXTH-GRADE CLASSES

	Criteria	N	Mean Generality of Pupils' Sociometric-Status Scores	Mean Generality of Teachers' Judgments	Differences
Boys	Seating-Play	40	.80	.68	-.12*
	Play-Work	40	.76	.63	-.13*
	Seating-Work	40	.86	.74	-.12*
Girls	Seating-Play	40	.76	.69	-.07†
	Play-Work	40	.76	.69	-.07†
	Seating-Work	40	.89	.75	-.14*

* Significant at the 1 per cent level.

† Significant at the 5 per cent level.

TABLE IV—CORRELATION COEFFICIENTS REPRESENTING THE CONSISTENCY OF THE EXTENT TO WHICH TEACHERS UNDER-GENERALIZED OR OVER-GENERALIZED IN THEIR JUDGMENTS OF SOCIOMETRIC STATUS OVER THE SEVERAL PAIRS OF CRITERIA IN FORTY SIXTH-GRADE CLASSES

	Boys		Girls	
	Play-Work	Seating-Work	Play-Work	Seating-Work
Seating-Play	.64*	.39†	.59*	.41*
Play-Work		.48*		.46*

* Significant at the 1 per cent level.

† Significant at the 5 per cent level.

teachers under-generalized in their judgments over the several criteria, while only ten of the teachers over-generalized. To determine if there was any consistency in the amount and direction with which teachers erred in generalizing over the criteria, the errors of generalization over the several pairs of criteria were correlated. It will be noted in Table IV that the extent to which teachers erred in generalizing on one pair of criteria was significantly related to their errors of generalization on the other pairs of criteria. In other words, the teachers' judgments indicated a fairly consistent tendency to consider sociometric status as being more specific to each criterion than it actually was.

DISCUSSION OF RESULTS

The correlation coefficients representing the generality of the accuracy of teachers' sociometric perceptions over criteria, in all probability, under-estimate the amount of generality present. The variations among teachers in their opportunity to observe choice behavior related to the play-companion criterion, the variations among classes and within classes in the spread of sociometric-status scores, and the general tendency for teachers to consider sociometric status as being more specific to each criterion than it actually was, all tended to reduce the indices of generality reported in this study. However, despite the influence of these variables, the correlation coefficients representing generality of the teachers' judgment accuracy were all found to be statistically significant. Thus it appears safe to conclude that the accuracy of teachers' sociometric perceptions has generality over sociometric criteria.

REFERENCES

- 1) D. P. Ausubel, H. M. Schiff, and E. B. Gasser, "A Preliminary Study of Developmental Trends in Sociempathy: Accuracy of Perception of Own and Others' Sociometric Status." *Child Development*, 1952, 23: 111-128.
- 2) I. E. Bender, and A. H. Hastorf, "The Perception of Persons: Forecasting Another Person's Responses on Three Personality Scales." *Journal of Abnormal and Social Psychology*, 1950, 45: 556-561.
- 3) N. L. Gage, "Judging Interests From Expressive Behavior." *Psychological Monographs*, 1952, 66, No. 18. (Whole No. 350), 20 pp.
- 4) N. E. Gronlund, "The Accuracy of Teachers' Judgments Concerning the Sociometric Status of Sixth-Grade Pupils." *Sociometry Monographs*, No. 25, 1951, 62 pp.
- 5) N. E. Gronlund, "The Relative Ability of Home-Room Teachers and Special-Subject Teachers to Judge the Social Acceptability of Pre-Adolescent Pupils." *Journal of Educational Research*, 1955, 48: 381-391.
- 6) R. M. W. Travers, "A Study in Judging the Opinions of Groups." *Archives of Psychology*, No. 266, 1941, 73 pp.
- 7) R. M. W. Travers, "The General Ability to Judge Group Knowledge." *The American Journal of Psychology*, 1943, 56: 95-99.

GROUP FUNCTIONING UNDER NEGATIVE CONDITIONS¹

LEONARD A. OSTLUND

Oklahoma A. and M. College

THE PROBLEM

The problem of the present study was to observe the behavior of a class of college students in an experimentally induced negative atmosphere. It was believed that a tension situation would provide a critical test of group functioning. The specific hypotheses tested were these:

- 1) The group will demonstrate a high degree of integration.
- 2) The group will demonstrate a high degree of adequate performance.

The students had been taught, for one semester, by the case-discussion method. The group was composed of three women and twenty-five men students, three-fourths of whom were juniors, the remainder seniors.

In a case-discussion course conducted in a permissive atmosphere created by the instructor, certain psychological benefits may accrue. These include the emergence of informal leaders, responsible behavior on the part of the students, and a fair degree of acceptance of each student by the group. Previous research by the writer had led to the conclusion that it would be appropriate to determine experimentally the effects of a negative atmosphere upon a case-discussion class as this might provide a further clue to some basic processes (2, 3, 4). The case used was "The Gordon Company" (1).

THE EXPERIMENTAL DESIGN AND RATIONALE

It was believed that the behavior of the students, with the foregoing conditions of negative atmosphere established, would be largely a function of the degree of group integration. In a previous research upon this same group, the writer demonstrated that it possessed a high degree of integration (4). Therefore, in the present experiment, it was predicted that this integration would persist

¹ The coöperation of Professor Edward G. Nelson and Dean William R. Butler, of the University of Kansas, facilitated the project.

and that the group would perform adequately in this experimentally created negative atmosphere. The latter was induced by confronting the students with a novel situation, namely, an attack upon them as a group by the writer, their instructor, who accused them of not having prepared the assignment and of not discussing the case in a responsible manner. The element of anxiety was heightened by involving the student's grade-point standing. This was a six-credit course, and the writer stated that he intended to inform the instructor who graded the group of the class's inadequate behavior.

The experimentally created negative atmosphere met criteria of "Lebensnähe", appropriateness, and spontaneity as is indicated below.

A) "*Lebensnähe*"—The writer believed that criticism which would have negative implications for grade-point standing would meet this criterion. His criticisms began by alluding to the lack of social responsibility shown by the class. At no time was any individual attacked. This was made clear by statements such as, "I want it to be clear that this is a question of group responsibility, no individual or several individuals are being singled out." The emotional tone of the responses indicated that the situation was real—the students acted seriously, not in fun. The atmosphere seemed charged with tension which mounted as the period progressed. Moreover, the transcript offered objective evidence in the form of fifteen pauses and this was considered indicative of tension. The words used by the students in explaining their behavior constituted a further indication that the matter was taken seriously.

B) *Appropriateness*.—The propitious situation arose in the following manner: The college was in an uproar because its basketball team was competing in the national championship and as a result, studies had suffered. At the previous session, the class disclosed that it had not prepared the assignment. This situation was used to justify the experimentally induced negative session at the next meeting.

C) *Spontaneity*.—The writer had been careful in his meetings with the class to maintain a permissive atmosphere. He already had the rapport of the students for he had acted as go-between for them in communicating their criticisms to the staff. When he suddenly criticized the group, the puzzled and bewildered expres-

sions and remarks of the students left no doubt that frustration was acute.

At the close of the period, the writer attempted to heal any psychological trauma and forestall an eruption in the next class. He explained that he had been critical purposely because he felt that this could prove a valuable experience for the students and the students' work had not been up to reasonable standards.

THE MEASURING INSTRUMENTS

In order to test the hypotheses stated on page 32, it was believed necessary to perceive the situation from the viewpoint of the subjects as well as from the viewpoint of neutral observers. This was done by means of students' ratings which compared the experimental period with others, and by judges' ratings of a verbatim transcript of the experimental period.

A) *Student opinion form.*—For several weeks, a form entitled "Student Opinion of Class Discussion" had been distributed at the conclusion of each period. Students were asked to rate the seven aspects of class discussion which appear in Table I. Opinions were rated in cells entitled "much less than usual," "less than usual," "same as usual," "more than usual," and "much more than usual." This form had been used many times in order to gain information concerning the students' opinions which could be used to enhance instruction. This made it possible to use these student judgments as a basis for comparing their feelings about and perceptions of the "negative" session with the pre-negative and post-negative class sessions.

TABLE I—CHI SQUARE RATINGS OF STUDENT OPINIONS, NEGATIVE VS. PRE- AND POST-NEGATIVE SESSIONS

Degree to which:	Pre-Negative	Post-Negative
1. You felt at ease	Not sig.	$P < .01$
2. You made progress	Not sig.	$P < .01$
3. Teacher assisted progress	$P < .01$	$P < .01$
4. Class contributed	Not sig.	$P < .03$
5. Class felt at ease	$P < .01$	$P < .01$
6. You contributed	Not sig.	Not sig.
7. Class made progress	Not sig.	$P < .01$

B) *Judges' ratings*.—Judges' ratings of the discussion that took place during the experimental class period were feasible because the entire session, which lasted forty minutes, was tape-recorded and subsequently typewritten with each speaker identified. The students had become accustomed to having their discussion recorded inasmuch as the writer had recorded every session for several weeks. A pilot study conducted at another college to test the judgment process led to modifications in the original plan. As a preliminary to the final judgment process, the writer held a briefing session with a new group of judges. He explained that the judgments were to be recorded on calibrated continua with "high" and "low" at either extreme and representing scores of zero and one hundred respectively. The judges were instructed to bear in mind the rationale of the two ratings to be made while they read the transcript and then to record their judgments and to cite statements in support of them. This was done in order to make the basis of judgment more concrete. The judges were instructed not to discuss the matter with anyone. The writer did not tell them that he had conducted the experiment, but explained that he was co-operating with others in having the transcript judged.

Twenty judges participated and completed the task within one month.² Judgments were recorded in a continuum or scale which ranged from zero to one hundred. The judges were given the following definitions: Group integration—"A high score indicates unity, that the class "stuck together," that it defended and supported the members as well as the group as a whole. A low score indicates a lack of unity, that the class "fell apart," individuals and the class as a whole were attacked or blamed." Group performance—"A high score indicates that the group tried to deal with the classroom problem by finding out what was wrong, as well as taking remedial action. A low score indicates that the group did not try to diagnose the situation and that it did not take corrective measures."

² The writer is indebted to the following judges: Members of his seminar "Psychology of Organized Groups": Doris Beard, Marilyn Born, Bob Brandeberry, J. L. Bryan, Warren Hendricks, Robin Hood, Jerry Johns, Vernon Kiker, Scott McDonald, Doyle Monger, Patricia Parnell, Perry Stinson and Joe Tidrow; Members of Psi Chi: Vernon H. Edmond, Eric Hemphill, Warren McClintock, Robert M. Morgan, Mary Seago, Jack Shelton and Jay Thomas.

ANALYSIS AND INTERPRETATION

A) *Student opinion form*.—The form, "Student Opinion of Class Discussion" was used to compare ratings of the pre-negative with the negative session by means of Chi Square. Differences were statistically significant ($P < .01$) on two ratings, "degree to which the teacher helped the class make progress" and "degree to which the class felt at ease." When the negative session was compared to the post-negative session, only one rating was not statistically significant—that entitled, "degree to which you contributed." The rating, "degree to which the class contributed" was significant at the 3 % level and the remainder were significant at the 1 % level. The data appear in Table I.

It seemed necessary to interpret the finding that at the pre-negative session only two variables had been rated as significantly different from the negative session, but that six out of seven variables were so rated at the post-negative session. Actually, the writer played his rôle of instructor the same way in both pre-negative and post-negative sessions. Therefore, it was concluded that these differences in ratings mirrored a more intensive effort by the class as well as a compensatory positive reaction which functioned to restore deflated egos. Thus analysis of this form afforded further support to the writer's observations and the transcript analysis that the negative session had been perceived as real.

B) *The transcript judgment analysis*. (1) *Integration*: Reference to Table II indicates that the tabulation of the *I* (integration) scores yielded a mean of 69.55, a median of 78.5, and a range from 23 to 95. The majority judged the group above 50 % in integration.

It will be remembered that judgments were supported by explanatory statements. Those who favored integration stated that students did not contradict each other and frequently used the terms "we" and "us." Moreover, attention was called to statements in which individuals defended other members of the group and challenged the instructor. Four judges rated integration below 50 % and gave as reasons bickering, individual-centered statements, and ineffective responses.

(2) *Performance*: Tabulation of the *P* (performance) scores (Table II) yielded a mean of 38.65, a median of 28.5, and a range from 15 to 90. Thus the majority rated performance below 50 %.

The majority stated that the responses of the students were repetitious, ineffective, and unorganized, and indicated that the

TABLE II—JUDGES' RATINGS

Judge No.	Integration	Performance	Diff.
*1	88	80	-8
*2	90	90	0
*3	27	24	-3
4	91	15	-76
5	85	30	-55
6	81	16	-65
7	70	44	-26
*8	80	55	-25
9	65	15	-50
10	86	24	-62
11	75	45	-30
12	75	25	-50
13	87	45	-42
14	50	30	-20
*15	32	26	-6
*16	34	19	-15
*17	95	75	-20
*18	77	64	-13
19	80	27	-53
*20	23	24	+1
\bar{X}	69.55	38.65	30.90
Median	78.50	28.50	50
Range	23-95	15-90	—

* Indicates possible "halo" effects.

class had accepted criticism passively. Five judges rated performance above 50 % and stated that the group had attempted to clarify, correct, and cope with the problem. The probable reasons for these discrepancies were that despite a common orientation, the judges had different frames of reference, and the data were difficult to evaluate.

(3) *Comparison of I vs. P ratings:* In all but two cases, performance was rated lower than integration. Even the anomalies were not far divergent for they involved differences between *I* and *P* scores of zero and plus one. The mean *I* score was 69.55, the mean *P* score was 38.65, and the mean difference score was 30.90. Analysis by a "t" test indicated that the difference was statistically significant ($P. < .01$). According to the judges, group integration was significantly higher than group performance.

Because there was a wide range which included extreme scores, the medians were examined. They were: Integration, 78.50, and Performance, 28.50. Thus a comparison of medians also suggests the substantial differences between judges' ratings of these two variables.

When the ratings were considered independently of each other, there were twenty *I* ratings of which four fell below 50% and twenty *P* ratings with five above 50%. However, independent analysis masked an important factor, namely that the ratings were made by the same judge. Therefore, the possibility of "halo" effects was studied next.

Inspection of the starred ratings of Table I revealed that in nine cases both ratings were either high or low. Further evidence of relationships between ratings resulted when the judges were separated into those who had rated both *I* and *P* on the same end of the continuum and those who had not. The mean difference for the nine judges whose ratings seemed related was ten points, whereas the mean difference for the remaining eleven judges was a spread of forty points.

SUMMARY

This experiment was designed to test two hypotheses related to the performance of a well-integrated class group under conditions of stress. Hypothesis I predicted that the class would demonstrate a high degree of group integration. Hypothesis II predicted that the class would demonstrate a high degree of adequate performance. In order to test these hypotheses, a case discussion class, previously demonstrated to be well-integrated, was presented with an experimentally produced negative atmosphere. The discussion throughout the experimental class period was tape recorded and rated by twenty judges in respect both to integration and performance. At the end of the experimental class period the students rated their feelings and perceptions of what had gone on, using a form that was regularly used at the end of each class period. This made it possible to compare students' feelings and perceptions about the experimental class with similar reactions to other classes.

The first hypothesis, that under conditions of stress this class would demonstrate a high degree of group integration, was supported by the judges' ratings. The second hypothesis, that the class would demonstrate a high degree of adequate performance,

was not supported by the ratings of the judges. The student ratings did not clearly support either hypothesis.

DISCUSSION

Several reasons occurred to the writer for the refutation of the hypothesis that the group would perform adequately even under conditions of stress. It is conceivable that the group's energies were expended in marshaling its forces, and that time ran out before progress in performance could demonstrate itself. It might be, too, that the attack by the instructor was so energetic and unexpected that the group was unable to recover its equilibrium.

The findings from this study seem to the author to have implications for the current controversy about the relative values of directive versus non-directive teaching. This class had been taught in a manner that was largely non-directive and it seemed to be able to maintain its integrity as a group, at least, quite successfully under conditions of stress. The writer believes, on the basis of a number of experiments as well as his own intimate, personal experience in case discussion teaching, that the most effective instructional procedure involves a judicious combination of both the directive and the non-directive types of teaching.

REFERENCES

- 1) J. D. Glover and R. M. Hower, *The Administrator*, Chicago, Richard D. Irwin, 1949.
- 2) Leonard A. Ostlund, "A sociometric matrix," *Proceedings of the Oklahoma Academy of Science*, Vol. 34, 1953, pp. 189-190.
- 3) Leonard A. Ostlund, "An experimental study of case discussion learning," *Journal of Experimental Education*, in press.
- 4) Leonard A. Ostlund, "Group integration in a case discussion course," *Journal of Educational Psychology*, Vol. 34, No. 8, December, 1953, pp. 463-474.

FACT AND PHANTASY IN VOCATIONAL INTEREST MEASUREMENT

HILDEGARDE JOHNSON

Iowa State College

A recent issue of Fortune magazine carries an article (1) in which a layman lays bare some very obvious fallacies in personality testing. It is time that researchers in the area of psychological testing do some very straight thinking and sound empirical research with respect to the validity of instruments they are placing on the market. Interest inventories might be just as subject to a layman attack as personality tests.

What is a valid scoring key for an occupation on an interest inventory? Apparently many users of interest inventories think that it is any scoring key for which norms have been reported or for which there is a significant difference between the means of an occupational and a base group.

Before developing a sound rationale for measuring validity it is necessary to consider the purpose of an occupational scoring key. A scoring key for an interest inventory is developed to measure those interests and aversions *peculiar* to persons employed in the occupation but not characteristic of persons employed in other occupations. In other words, an attempt has been made to differentiate the interests of two populations: persons employed in an occupation and persons not employed in it by excluding those interests and aversions which they have in common. Therefore, it seems reasonable that a scoring key for an interest inventory has satisfactory concurrent validity if it widely separates these two groups on the basis of the distributions of scores of a sample of persons.

If validity should be expressed in terms of the success of separation of two distributions of scores on the occupational scoring key, how may this separation be measured? Obviously, some measure of the amount of overlapping between the two distributions is indicated. This might be shown by these four methods (1) percentage of the area of the two curves which lies in the overlapping region, (2) percentage of cases in one distribution which may be matched with cases in the other distribution, (3) percentage of cases in the

reference group which exceed the mean of the occupational group, and (4) mean difference expressed in standard deviation units. The first of these is probably the most easily interpreted for most persons who might use the Inventory.

A standard for measuring validity is needed before numerical expressions of amount of overlapping can be interpreted. Strong (2) reasoned that since persons employed in an occupation have some interests in common with persons employed in other occupations the overlapping which actually exists, not zero percent overlapping, is this standard for evaluating validity. Such a criterion would always be an unknown quantity. This varying and unknown criterion of validity must be accepted if it is true (1) that each scoring key measures interest in one occupation and (2) that comparison of interest in a large number of pairs of occupations is of primary concern. However, each key is not a measure of all interests and aversions characteristic of persons employed in the occupation but only of those *peculiar* to them and *not characteristic* of persons in a reference group. An attempt is made to exclude interests and aversions common to the two groups. Therefore, point one seems illogical.

With respect to point two, the difficulty of measuring and interpreting validity in terms of overlapping between distributions of occupational groups taken two at a time will be apparent if we consider just ten occupations. The combination of these ten taken two at a time is equal to 45. If overlapping was computed twice for each pair, once for each of the occupational scoring keys, there would be 90 measures of validity for the ten occupational scoring keys, none of which could be compared with a known standard of validity. When using inventory scores in guidance, the problem is not one of comparing a student's interest in these 45 pairs of occupations. It is a problem of comparing his interest in each occupation with the interest of that reference group to which the student logically belongs.

Since it is not true that each scoring key measures interest in one occupation and that a comparison of interest in pairs of occupations is of primary concern it seems reasonable to reject Strong's reasoning concerning a standard for use in interpreting percentage of overlapping. Interests peculiar to persons employed in the occupation but not characteristic of persons in the reference group are identified and measured. If this objective is perfectly achieved,

the percentage of overlapping between the distributions of scores of persons employed in the occupation and of persons in the reference group will be zero. Thus, there is a standard for interpreting the measure of validity for each occupational scoring key, zero overlapping with the reference group.

There is little likelihood that the distributions of an occupational and reference group will not overlap at all, just as an achievement test will seldom, if ever, correlate perfectly with the criterion used in determining its validity. Validity of the various scoring keys for an interest inventory may, however, be compared.

Studies of concurrent validity should be cross-validation studies in which the occupational and reference groups are made up of persons not included in the samples used for developing scoring keys. Before it has been possible to secure these new samples, expressions of validity based on the original samples are much better than publication of the test with no statement of percentage of overlapping between criterion and reference groups.

Predictive validity of occupational keys for an interest inventory is of greater importance than concurrent validity. Therefore, it is necessary to use a longitudinal research design in which persons respond to the inventory when they are students and again a number of years later when they are employed. Inventory scores are used to predict that a student will or will not find satisfaction in certain occupations, not that he will or will not enter each of the occupations. Therefore, the hypothesis should then be tested that persons achieve greater vocational satisfaction when they enter an occupation in which their interest inventory scores were high enough for this occupation to be recommended for consideration. Testing this hypothesis presupposes an adequate measure of vocational satisfaction.

Strong (2) studied predictive as well as concurrent validity of the Vocational Interest Blank for Men. A ten-year follow-up was made of college seniors and a nine-year follow-up of college freshmen. Since no adequate measure of vocational satisfaction was available, Strong (2), used continuance in an occupation as a measure of satisfactory adjustment.

Measures of validity are not reported for some widely used interest inventories. Only the scoring keys and norms are made available. The consumers of the test are supposed to assume that the test measures interest peculiar to each occupational group for

which there is a scoring key and that if a student scores as high as, say the thirtieth percentile, his interests are peculiarly like those of persons employed in the occupation. That this phantasy is frequently not true may be inferred by studying Table 1. The low validity (high percentage of overlapping) of certain of these scoring keys is startling.

When data on overlapping are available, these should be used both in choosing the inventory to use for vocational guidance and in counseling with students. Suppose that the percentage of overlapping between scores of an occupational group and the reference group is 60. Suppose, also, that we ignore this information when interpreting the scores of students. The interests of a student scoring at the mean of the norm group might certainly be expected to compare favorably with those of persons employed in this occupation. This student scores at the fiftieth percentile. Usually, the counselor would tell him that on the basis of interests, this is an occupation which he should consider seriously. Considering the 60 per cent overlap, however, 14.8 per cent of the persons in the reference group score higher than this student. If we classify him in the occupational group, there is a 15-in-100 chance that he actually belongs in the reference group.

Suppose, next, that the percentage of overlapping is 25 and that a student scores at the mean of the criterion group. If we say that, on the basis of interests measured by this test, the student belongs with the occupational group, there is a one-in-100 chance that he has been incorrectly classified.

In the ideal situation, each time a counselor recommends that a student consider a certain occupation, the student should be told the probability of error, or the probability that he really belongs in the reference group with respect to this choice. This ideal would be difficult to achieve.

In order that information about percentage of overlapping could be used by the counselor, arbitrary cutting points might be located on the norms for each occupation, say the ninety-fifth percentile of the reference group. When using the norms in guidance it would then be possible to say with a five-in-100 chance of error that a student scoring beyond this point belongs in the occupational group. Unfortunately this procedure has not been used for interest inventories which are on the market. However, information concerning the percentage of overlapping with the reference group is

TABLE 1. OVERLAPPING BETWEEN OCCUPATIONAL AND REFERENCE GROUPS OF SCORES ON SELECTED SCALES OF INTEREST INVENTORIES

Investigator	Inventory	Scale	Occupational Group	N	Reference Group	N	% Overlap.	% = or > Mean or Median
Clark (3)	Minn. Voc. Interest Inventory	Painters	Painters	252	Tradesmen-in-gen.	175	30.5*	2.0
		Sheet metal workers	Sheet metal workers	99	Tradesmen-in-gen.	175	59.4*	14.3
			Median of 8 groups	113	Tradesmen-in-gen.	175	44.5*	6.3
Kriedt (4)	Voc. Interest Blank for Men	Guidance psych.	Guidance psych.	115	Psych.-in-gen.	933	49.6*	8.7
		Clinical psych.	Clinical psych.	221	Psych.-in-gen.	827	60.3*	14.9
Strong and Tucker (5)	Voc. Interest Blank for Men and Med. Specialists Preference Blank	Internist	Internist	100	Physicians-in-gen.	100	48.0	7.9*
		Surgeon	Surgeon	100	Physicians-in-gen.	100	42.0	5.3*
		Pathologist	Pathologist	100	Physicians-in-gen.	100	37.0	3.6*
		Psychiatrist	Psychiatrist	100	Physicians-in-gen.	100	41.0	5.0*
Strong (6)	Voc. Interest Blank for Women	Home econ. teacher	Home econ. teachers		Women-in-gen.	135	32.0*	2.3*
		Dietitians	Dietitians		Women-in-gen.	135	35.0*	3.1*

Lehman (7)	Kuder Pref. Record	Artistic	Home econ. teachers	125	Home econ. in-gen.	218	89.3*	39.4*
		Clerical	Home econ. teachers	74	Home econ. in-gen.	189	84.6*	35.0*
		Computational	Home econ. teachers	125	Home econ. in-gen.	218	98.7*	48.7*
		Literary	Home econ. teachers	125	Home econ. in-gen.	218	89.6*	39.7*
		Mechanical	Home econ. teachers	74	Home econ. in-gen.	189	99.4*	49.4*
		Musical	Home econ. teachers	125	Home econ. in-gen.	218	98.3*	48.3*
		Persuasive	Home econ. teachers	125	Home econ. in-gen.	218	78.4*	29.1*
		Scientific	Home econ. teachers	125	Home econ. in-gen.	218	96.3*	46.3*
		Soc. Service	Home econ. teachers	125	Home econ. in-gen.	218	93.6*	43.6*

* Numbers supplied from Tilton's table by referring to $\frac{Mn_i - Mn}{SD}$, or % = or > Mn. J. W. Tilton. The measurement of overlapping. *Journal of Educational Psychology*, 28: 656-662. 1937. The interested reader may obtain confirmation of these computations by writing to Hildegard Johnson, Division of Home Economics, Iowa State College, Ames, Iowa.

sometimes available in research reports in professional journals. Such information should help the counsellor to know when his recommendation that a student consider a certain occupation should be presented as a recommendation in which there is considerable chance of error and when it should be presented as a recommendation in which the student may have considerable confidence.

The consequence of using, in guidance, inventories which do not successfully differentiate occupational groups from persons who are not employed in them or which do not predict vocational satisfaction is that we have more "square pegs" in "round holes." Vocational "misinformation" may be as serious as the diagnosis of a quack doctor. The emphasis which has been placed on the student's enjoyment in taking the test, ease of scoring and agreement with expressed interests clouds the real issue. Emphasis needs to be placed upon selection of inventories which measure what they are supposed to measure.

In conclusion, a reasonable test of the validity of an interest inventory can be made. This same test, percentage of overlapping between criterion and reference groups, can then be used as a basis for supplying counselors with some information about the probability that a student belongs in the reference group rather than the occupational group. Recommendations based on such information would be sound and would lead to improved vocational guidance.

REFERENCES

- 1) William Whyte, Jr, "The fallacies of personality testing." *Fortune* L No. 3, 117-119, 204-204, 1954.
- 2) E. K. Strong, *Vocational interests of men and women*. Stanford University, California, Stanford University Press, 1943.
- 3) Kenneth E. Clark, A vocational interest test at the skilled trades level. *Journal of Applied Psychology*, 33: 298-299, 1949.
- 4) Phillip H. Kriedt, Vocational interests of psychologists. *Journal of Applied Psychology* 33: 484, 486, 1949.
- 5) Edward K. Strong and Anthony C. Tucker, The use of vocational interest scales in planning a medical career. *Psychological Monographs*. Vol. 66. no. 9, 1952.
- 6) Edward K. Strong, Information on the distribution of women-in-general on the dietitian and home economics teacher scales (Private communication) 1950.
- 7) Ruth T. Lehman, Raw scores of practicing home economists on the Kuder test (Private communication) 1949.

READING IMPROVEMENT AS A FUNCTION OF STUDENT PERSONALITY AND TEACHING METHOD

DONALD E. P. SMITH, ROGER L. WOOD, JAMES W. DOWNER
and ALTON L. RAYGOR

University of Michigan

The nature of teaching method-learning product interaction has long been a problem stimulating research (2, 19). The weight of evidence appears to favor Wolfle's conclusion that teaching method is a relatively unimportant variable. Recent evidence on learner personality-method interaction (13, 18), however, indicates that the problem is yet to be settled unequivocally. As a result of such evidence, Birney and McKeachie have suggested that, in teaching-methods research, designs should be utilized which remove the variance attributable to relatively permanent personality traits (2).

This paper reports results of the gains in reading efficiency among college students when the teaching method was varied systematically for groups of subjects representative of two personality syndromes.

PRELIMINARY WORK

The hypotheses tested in the present study are derived from three lines of investigation: (1) personality structure; (2) course organization and anxiety; and (3) personality structure and reading improvement.

(1) *Personality Structure*. The first line of investigation resulted in a partial synthesis of Freudian and Lewinian views on personality structure, anxiety and the effects of anxiety within individuals differing in structure (12). The senior writer suggested that two important dimensions of personality are: (1) a generalized anxiety and (2) degree of permeability.¹ The second dimension is conceived as relative thickness of "tension system" boundaries, in Lewin's terms (8), giving rise to differences in "permeability" or ease of communication among systems. It was further suggested that

¹ Originally termed flexibility-rigidity. The term was abandoned due to the many conflicting conceptions of rigidity (1).

anxiety is relatively unrelated to permeability with the result that we should expect to find individuals with varying degrees of each in at least two dimensional space.

Combinations of extreme degrees of both dimensions might then appear as in Figure 1.

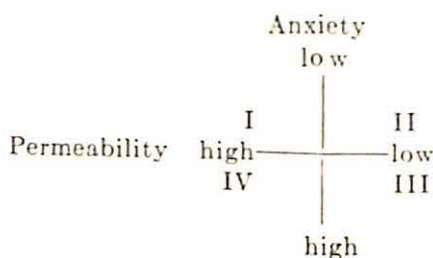


FIG. 1

Type I. Individuals who are highly permeable and low in anxiety tend to be changeable, sociable, impulsive, sensitive, imaginative and stable.

Type II. Those who are relatively impermeable and low in anxiety tend to be conservative, independent, conscientious, somewhat unsociable and stable.

Type III. Like II, these individuals are conscientious but very withdrawn, aggressively independent, suspicious, rigid in attitudes and, when in extreme disequilibrium, given to paranoid delusions.

Type IV. The combination of anxiety and permeability results in individuals who are disorganized, sensitive, especially in interpersonal relations, imaginative, excessively dependent and cycloid.

Descriptive categories, similar to IV and III have been published by others (6, 18). Frank *et al.*, have termed individuals of one pattern the "doctor's assistant" and the other the "help-rejecting complainer." Wispé describes two types as "insecure" and "independent."

(2) *Course Organization and Anxiety.* The second line of investigation was an attempt to determine the effect of anxiety in the learning situation on "applicational transfer" of the content of an introductory psychology course (13, 14). One result of the study was evidence that students differ markedly in anxiety as a function of teaching method. Despite instructor behaviors and class procedures meant to induce anxiety, some twenty per cent of the students reported complete satisfaction with the conduct of the course. On the other hand, classes characterized by minimal structure and

considerable permissiveness included a like percentage of students who reported dissatisfaction and anxiety attributable to the learning experience.

It appeared that some individuals felt secure and happy in a highly structured class with an autocratic, even punitive, instructor, but were insecure in a permissive class with a warm, accepting instructor.

(3) *Personality Structure and Reading Improvement.* The third line of investigation involved determination of some relationships between personality structure and reading improvement in a university setting, reported elsewhere (15). In brief, groups of students ($N = 18$) improving very slowly or very quickly in their reading efficiency responded to Cattell's Sixteen Personality Factor Test (4). *Tau* coefficients of correlation among the resulting profiles were determined for every *S* with every other *S*. The resulting matrix was analyzed by inspection and clusters of *Ss* similar to each other but different from other clusters were drawn. Four clusters were found, the mean profiles of two of which were similar in anxiety but different in sociability and "independent self-sufficiency" (4, p. 11). In general, the first cluster seemed partly described by Cattell's term "cyclothymic" and partly by the traits "irresolute" and "fickle"; the second is described by the term "schizothymic" and by factor *Q2*, "strong willed." The other two clusters paralleled the first two but were differentiated primarily by the anxiety factors. Replication with a larger sample ($N = 25$) confirmed the initial findings.

Interestingly, the several trait names provided by the Sixteen Personality Factor manual for the four clusters are remarkably similar to those used earlier to differentiate extreme categories on the permeability-anxiety dimensions. For the purpose of the present orientation, permeability and anxiety will be defined in terms of the above clusters. The two groups similar to each other in anxiety were pooled and compared with the less anxious groups. The two groups similar at the cycloid extreme were pooled and compared with the two at the schizoid extreme. Two measures of approximately forty items each were thus derived from the three hundred and seventy-four items constituting forms A and B of the 16 PF inventory. The criterion of selection was a response differential significant at the 0.01 level.

The two measures were administered to a group of general psy-

chology students ($N = 100$) and an internal consistency procedure was carried out. After discarding items with r s below 0.20, thirty-three items remained in the permeability scale and thirty-two in the anxiety scale.² Corrected split-half coefficients of reliability are: $r_{\text{perm.}} = 0.64$; $r_{\text{anx.}} = 0.81$. [The $r_{\text{perm.}}$ is a "serious underestimate" according to Guttman's criterion of dissimilar σ s among the halves of the scale (7).]

A number of studies have been carried out in order to determine the validity of the scale in terms of external criteria. One of these, a cross validation procedure, yielded an r of 0.46 between permeability and gain in reading efficiency with anxiety constant ($P < 0.01$). A second study indicated a mean gain in reading efficiency of types I and IV roughly one hundred per cent greater than that of types II and III ($N = 30$).

HYPOTHESES

The three lines of research, those on personality structure, on teaching method and anxiety, and on personality and reading improvement, have culminated in the following hypotheses:

(1) Type IV individuals, described as permeable or fluid, cycloid and disorganized, will feel most secure in a structured course with a "directive" instructor and, consequently, will make optimum gains in reading efficiency with such treatment, but will feel most insecure in an unstructured course with a permissive or "non-directive" instructor and will make minimal gains in reading efficiency with such treatment.

(2) Type III individuals, described as impermeable or rigid, schizoid and over-organized, will feel most secure, thus, most free to change existing structure, in an unstructured course with a permissive instructor and will make optimum gains in reading efficiency with such treatment, but will feel most insecure or threatened in a structured course with a "directive" instructor and will make minimal gains in reading efficiency with such treatment.

(3) Type IV individuals will profit more from individual instruction than from class participation as a result of (a) opportunity to develop a secure relationship, and (b) reduction in distractions or in the effect of group tensions.

² A copy of the scale including certain validating information is available from the Division of Reading Improvement Services, Bureau of Psychological Services, University of Michigan.

(4) Type III individuals will profit more from group participation than from individual instruction as a result of the anonymity afforded by group membership and a consequent reduction of threat from the instructor (viewed as an authority figure).

With respect to the final hypothesis, Miller cites Hutt to the effect that psychoneurotics who give constricted Rorschach records when taking the individual form give dilated records in the group where they seem to gain security from their relative anonymity (9, p. 368).

The present study reports the results of a test of the first two hypotheses.

INVESTIGATION

A balanced design with two treatments, four instructors and sixteen classes was carried out in order to control some of the many factors which might contribute to variance in the criterion measure. As will be shown, at least one of the apparent selective factors did not yield to control by this design so that an analysis of covariance was required.

SUBJECTS

Experimental Ss were eighty college students, fifty males and thirty females, selected from two hundred and twenty-five who chose to register for seven-week corrective-reading classes in a university reading service. Approximately half registered for the first series of classes and the remainder registered for the second series during the same semester. Ss were assigned to experimental classes on the basis of personality inventory scores placing them in one of the two high anxiety classifications described above, III or IV. Cut-off points in the selection procedure were: for anxiety, fiftieth percentile on reading-client norms (raw score of eighteen or below); for permeability, below the thirtieth percentile (raw of eleven) or above the seventieth percentile (raw of seventeen). Whereas Ss were similar in anxiety, they were quite dissimilar in permeability.

PROCEDURE

Ss were placed in classes on the basis of inventory scores only. An attempt was made to limit all classes to twelve members, five

TABLE I
Series I

Instructor	A		B		C		D	
Treatment	Directive		Directive		Non-directive		Non-directive	
Type and class order . . .	III	IV	IV	III	IV	III	III	IV
Prediction	—	+	+	—	—	+	+	—

Series II

Treatment	Non-directive		Non-Directive		Directive		Directive	
Type and class order	IV	III	III	IV	III	IV	IV	III
Prediction	—	+	+	—	—	+	+	—

experimental *Ss* and the remainder types I, II and those falling in the middle of the distributions of permeability scores.

Design. Each of four experienced instructors taught four classes, two in each series. Each used one treatment for both personality types during one seven-week series of classes, then changed treatments during the second series. The design appears in Table I. "Class order" refers to the time sequence of classes and "prediction" refers to maximum (+) or minimum (—) gain in reading efficiency.

External control variables consisted of age, sex, year in college, scholastic aptitude as inferred from scores on the *ACE Psychological Examination for College Freshman*, and initial reading rate (words per minute) and story comprehension as inferred from scores on form A of the "Survey" section of the *Diagnostic Reading Tests* (5). Since *Ss* were assigned to sections on the basis of personality inventory scores, it was assumed that mean differences in the control variables between types receiving each treatment would be negligible. The assumption was tested and confirmed in regard to age, year in college, scholastic aptitude, reading rate and sex. It is of interest to note that sixty-eight per cent of the males were categorized as Type III and thirty-two per cent as IV, whereas females were classified evenly, fifty per cent in each group.

One variable, initial reading comprehension, was significantly higher ($P < 0.01$) at the mean for Type III Ss receiving directive treatment than for III Ss receiving non-directive treatment.

Gain. The criterion of gain was the increase, on an equivalent form (B), in "rate of comprehension" (RC) expressed as a percentage of the initial RC. Rate of comprehension is the product of rate of reading and percentage of comprehension items answered correctly. Rate is based on reading selections of about one thousand six hundred words, and comprehension is inferred from responses to twenty multiple choice items sampling immediate retention of details and ability to infer implied meanings and to draw conclusions.

An evaluation form, described below, was completed anonymously by the Ss before they had received the results of the final reading test.

Treatment. The instructors became thoroughly familiar with the "directive" viewpoint toward counseling sometimes ascribed to Williamson (17) and Thorne (16), and to the "non-directive" theory derived from Rank and advocated by Rogers (10) and, perhaps, by Cantor (3). It was agreed that the teacher behaviors shown in Table II seem to be representative of the two extreme views when those views are applied to the teaching of corrective reading classes:

The tone was set, apparently, at the first class meeting. One instructor reported that he was unable to change the tone after the first two weeks. This will be discussed further in a later section.

Course Content and Facilities. Reading improvement procedures were similar to those used in most college programs. Students met in small groups for two one-hour sessions each week for seven weeks, and were given an opportunity to use the laboratory facilities at their convenience. The class meetings were devoted to a few or all of the following activities depending upon the treatment: instruction in the psychology of reading, study techniques and examination writing; discussion of problems relating to reading, study and exam panic; timed reading exercises; vocabulary and critical reading exercises; practice with pacing devices, a tachistoscope and reading films. A supply of mimeographed exercises was available to enable students to work on their own. Some "gorged" themselves on reading films, others on the pacers or tachistoscope.

TABLE II

Situation	Instructor Behaviors	
	Directive	Non-directive
Orientation		
Seating	"If easily distracted use small tables near wall. Keep to one place."	—
Work and progress records	"Keep complete"	"for your convenience"
Improvement plan	Instructor fills out a mimeographed form.	"for your convenience" Explain meaning of diagnostic results, appropriate procedures. Be available if <i>S</i> demands help.
Attendance	"Excused absence only"	Voluntary
Course structure	Explain operations	"seven weeks"
Initial interview	Compulsory	Voluntary
Lab. practice	"Determine practice time and enter on schedule."	"Lab will be open if you wish to practice."
Class Conduct		
Record keeping	Careful, point-by-point instructions.	Explain when <i>S</i> asks.
Activities information	Lectures	Non-directive discussion (3, 10).
Order of exercises	Provide schedule of activities before class.	Class determination.
Procedures	Standard	Individuals determine own activities.
Instructor's De-meanor	Formal, dominant, supportive. Introduce self by title.	Informal supportive. Introduce self by given name.

RESULTS

Differences in the Criterion. Gains in the criterion measure and rate of comprehension, are presented in Table III. The gains are expressed in terms of percentage of increase over initial rate of comprehension. Type IV individuals increased 91.4 per cent in efficiency under directive treatment and 46.0 per cent under non-directive treatment. The difference of 45.4 per cent is significant beyond 0.01. Type III individuals increased 56.0 per cent under directive

TABLE III.—COMPARATIVE INCREASE IN READING EFFICIENCY (RATE OF COMPREHENSION) OF DIFFERING PERSONALITY TYPES AS A FUNCTION OF TREATMENT METHODS

Treatment	N	IV	Predict.	N	III	Predict.
Directive	12	91.4%	+	19	56.0%	—
Non-directive	15	46.0%	—	23	94.8%	+
Diff.		45.4%			38.8%	
<i>t</i>		3.30*			6.12*	

* $P < 0.01$

treatment and 94.8 per cent under non-directive treatment. Again the difference of 38.8 per cent is significant beyond 0.01.

While the above results apparently confirm both hypotheses, one variable has been left uncontrolled. It was found that both of the Type IV groups and both of the Type III groups were equivalent in terms of age, year in college, scholastic aptitude and initial rate of reading. The Type III groups differed, however, in initial reading comprehension, the group receiving non-directive treatment having a significantly lower initial comprehension than that receiving directive treatment.

Since differences in gain might be attributable to initial comprehension, an analysis of covariance was utilized, and the variance contributed by initial comprehension was removed. Results are shown in Table IV.

It appears that the difference in gain between the Type III groups is attributable to differences in initial comprehension rather than to varying treatments, as indicated by the resulting F of 0.10. It is possible that the large increase at the mean of those

TABLE IV.—SIGNIFICANCE OF DIFFERENCES IN READING GAIN OF TWO PERSONALITY TYPES RECEIVING DIFFERING TREATMENTS, WITH VARIANCE ATTRIBUTABLE TO INITIAL READING COMPREHENSION REMOVED

Source of Variation	III		IV	
	<i>df</i>	$V_{x,y}$ *	<i>df</i>	$V_{x,y}$
Among means	1	263	1	9104
Within groups	39	2697	24	1061

 $F = 0.10$ NS $F = 8.58, P < 0.01$

* Variance of gain with initial reading comprehension removed.

receiving non-directive teaching results from a regression effect, i.e., those scoring low in initial comprehension may have a tendency to score higher on a retest irrespective of training.

While the difference in gain disappears for Type III groups, the difference remains significant among Type IV groups as indicated by an F of 8.58 ($P < 0.01$).

Instructor-method Interaction. No instructor-method interaction was found. This result must be accepted cautiously since the number of Ss in some groups was small. In sixteen classes, two had only two Ss, three had three Ss, two had four, seven had five, one had six and one, seven Ss.

Attrition. As is common in non-tuition, non-credit reading courses, the drop-out rate is relatively high, usually approaching twenty per cent of those who enroll. In the present study, total attrition among experimental Ss was sixteen per cent, virtually identical for III's and IV's. Similar numbers dropped from the III groups despite treatment. Among the IV's, one dropped from the directive class while three withdrew from the perhaps more threatening non-directive class.

Absences. Mean number of absences for Type III Ss was 3.0 in the permissive classes and 1.0 in the directive, in line with the greater independence and conscientiousness characteristic of those individuals. Type IV Ss were absent 1.9 meetings in directive classes and 1.7 meetings in permissive classes, an interesting reversal. Dependence characterizes this group.

Subjective Evaluation. Ss completed, anonymously, a questionnaire before receiving the results of the retest. Items included an evaluation of progress in reading, the effect on grades, the instructor's management of the class, Ss preference for group or individual activities, changes in ability to handle college work, and changes in ability to determine how well an assignment has been understood. Most of the differences between types and treatments were negligible. In general, IIIs prefer to work by themselves and IVs prefer group instruction as might be predicted. Type IIIs receiving directive treatment were more sure that their reading had improved than were those in the unstructured course ($P < 0.05$).

Instructors' Reactions. Although the instructors were usually unaware of the type of experimental Ss included in their classes, they were painfully aware of the method they were using. All reported considerable anxiety and frustration while using non-directive

tive techniques. One instructor exceeded his tolerance after two weeks in one class and attempted to teach normally, i.e., authoritatively, to his regret. His students ignored his directions and continued to work independently.

There was general agreement that those Ss receiving non-directive teaching were missing out on so much instruction and practice that they would probably make little progress. As it turned out, IIIs made as good progress (sixty-five per cent, corrected for initial comprehension) with such methods as did those with directive teaching (sixty-seven per cent), although IVs suffered as predicted. This constitutes some indirect confirmation of hypothesis two.

Finally, it was thought that seven weeks is probably too brief a period to expect individuals characterized as IIIs to blossom with non-directive teaching.

CONCLUSIONS

The first major conclusion of the present study derives from the confirmation of hypothesis one: anxious individuals of permeable or fluid structure will make optimum gains in reading efficiency when exposed to a maximum of course structure and direction, and will make minimum gains when little structure and direction is utilized. It may be concluded that, within the limits of the size and nature of the sample and of the validity of the criterion measures, the method of teaching is a major factor in the learning of individuals who are both anxious and fluid in structure.

The second major conclusion derives from the apparent failure to confirm hypothesis two: anxious individuals of impermeable, rigid structure will make optimum progress when exposed to non-directive teaching procedures and will make minimum progress when subjected to rigid class structure and considerable direction. Since progress was adequate (sixty-six per cent increase in reading efficiency) though less than that of directionally taught IVs (ninety-one per cent), it is concluded that either (1) the theoretical formulation is incorrect, or (2) experimental conditions were inadequate to test the hypothesis, or (3) the advantages of non-direction were offset by failure to use the proper facilities or to use them sufficiently.

There appears to be no evidence for assessing those alternatives other than the speculations of the instructors. A combination of the second alternative, specifically, inadequate length of treatment,

and of the third alternative, specifically, inadequate information and too many absences, seems likely, but cannot be accepted without further tests.

IMPLICATIONS

The above results seem to confirm the commonly held belief that teaching method is an important variable in the learning process, at least for some kinds of learners.

The now almost defunct dispute among directive and non-directive therapists (11, 16) also may be further clarified by the above results. Apparently more than a combination of techniques is required of the therapist. Some extreme individuals will, perhaps, thrive under one method or the other exclusively, while the normative individual may require some variation in treatment, determined by the ebb and flow of the treatment process. Progress appears to be a function of anxiety reduction, controlled to some extent by the therapist, and a resulting freedom for self direction.

Type IV individuals gain security from their dependence upon structure provided by an authority figure, and are then free to learn patterns of living similar to that structure, thereby reducing threat. Type III individuals gain security from their acceptance by the therapist, and are then free to change patterns of living to conform with their ideals.

SUMMARY

Several lines of research on personality structure, teaching methods and reading improvement as a function of personality culminated in a series of hypotheses concerning teaching methods-learning-product interaction with learner personality systematically controlled. Two dimensions of personality were described, a generalized anxiety and permeability or fluidity of structure, and inventory items purporting to sample the dimensions were listed.

With increase in reading efficiency as the criterion of learning and with a balanced design using four instructors and sixteen classes, college students were exposed to two methods of teaching: directive, structured; and non-directive, unstructured. Students characterized as permeable and anxious made optimum progress with directive methods. Those described as impermeable and anxious were, apparently, uninfluenced by the teaching method as inferred from gain in efficiency.

Implications for theoretical problems of teaching and counseling were discussed.

REFERENCES

- 1) Dee G. Applesweig, "Some Determinants of Behavioral Rigidity," *Journal of Abnormal and Social Psychology*, 49: 224-228, 1954.
- 2) R. Birney and W. McKeachie, "The Teaching of Psychology: A Survey of Research since 1942," (in press).
- 3) N. Cantor, *The Dynamics of Learning*. New York, Foster and Stewart, 1946.
- 4) R. B. Cattell, D. R. Saunders, and G. Stice, *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, Institute for Personality and Ability Testing, 1953.
- 5) Committee on Diagnostic Reading Tests, Inc., *Diagnostic Reading Tests: Directions*. Chicago, Science Research Associates, 1952.
- 6) J. Frank, et al., "Two Behavior Patterns in Therapeutic Groups and Their Apparent Motivation," *Human Relations*, 5: 289-317, 1952.
- 7) L. Guttman, "A Basis for Analyzing Test-retest Reliability," *Psychomet.*, 10: 255-82, 1945.
- 8) K. Lewin, *A Dynamic Theory of Personality*. New York, McGraw-Hill Book Co., 1935.
- 9) D. Miller, "Prediction of Behavior by Means of the Rorschach Test," *Journal of Abnormal and Social Psychology*, 48: 367-75, 1953.
- 10) C. Rogers, *Client-Centered Therapy*. New York, Houghton Mifflin, 1951.
- 11) W. U. Snyder, "Dr. Thorne's Critique of Nondirective Therapy," *Journal of Abnormal and Social Psychology*, 40: 336-339, 1945.
- 12) D. E. P. Smith, "A Stability-rigidity Theory of Personality Processes," (unpublished manuscript) 1951, (mime. copy available on request).
- 13) D. E. P. Smith, *Applicational Transfer: Its Nature, Measurement and Improvement*. PhD. thesis, Cornell University, 1952.
- 14) D. E. P. Smith, "Applicational Transfer and Inhibition," *Journal of Educational Psychology*, 45: 169-74, 1954.
- 15) D. E. P. Smith and R. L. Wood, "Noncognitive Personality Dimensions as Indices of Reading Improvement at the College Level," *American Psychologist*, 8: 438, 1953 (abstract).
- 16) F. D. Thorne, "A Critique of Nondirective Methods of Psychotherapy," *Journal of Abnormal and Social Psychology*, 39: 459-70, 1944.
- 17) E. G. Williamson, *How to Counsel Students*. New York, McGraw-Hill Book Co., 1939.
- 18) L. G. Wispé, "Evaluating Section Teaching Methods in the Introductory Course," *Journal of Educational Research*, 45: 161-86, 1951.
- 19) D. L. Wolfe, "The First Course in Psychology," *Psychological Bulletin*, 39: 685-712, 1942.

BOOK REVIEWS

PAUL V. LEMKAU. *Mental Hygiene in Public Health*, second edition. New York: McGraw-Hill, 1955, pp. xiii; 486.

This is a second and much improved edition of a book which first appeared in 1949. It deals with problems of mental hygiene, now more commonly called mental health. With wide experience and extensive study the author has attempted to deal with his subject from the point of view of the community and to answer in a practical way what he considers to be the most important and ever recurring questions.

Part One deals with the place of mental hygiene in public health and discusses in seven chapters the field of mental hygiene, the responsibility of the public, personnel techniques and concrete procedures, and the handling of problems from the points of view of national services, state organization and local administration.

Part Two discusses the development of the individual in chapters eight through eighteen. It deals with personality development, eugenics, and seven periods of life from the prenatal through old age.

An appendix gives a brief review of psychopathological states.

The preparation and mature experience of the author permit giving many valuable suggestions for the organization and administration of mental health work for any community which desires to set up and carry on this work, including answers to many questions that continually arise. The helps, visual aids, for example, which are available are indicated, and advice given for the use of newspapers, radio, posters, etc. If those interested in mental health want to know what can be done, they are told; if they want to know how to do it, they will find many usable suggestions. There is enough historical material to give some orientation in the development of the entire field; and there is detailed information as to the services that can be obtained from both national and state services.

The discussions of periods of mental development do not take the place of courses in child study and adolescence, etc., but they do present many facts and suggestions for dealing with problems of different ages which should be helpful for teachers and social workers of the different periods. Many of the differences in the problems found at different ages are pointed out.

The book is quite different from most of the texts on mental hygiene and the psychology of adjustment, but it is meant to be so. The emphasis is not that of handling problems from the point of view of the individual, but of the community. It is the work of a medical man, a Professor of Public Health Administration at the Johns Hopkins University, and of a Director of Mental Health Service on a community mental health board. Its appeal is to community workers in public health and the revised edition meets the needs of such workers in many practical ways.

The appendix, "Review of Psychopathological States," apparently has not been brought quite as fully up to date as has the main text.

The judgment is generally excellent and mature, and there is a sympathetic appreciation of the matters discussed. Workers in this field who study this volume cannot fail to learn the points of view of the other workers and should find it easier to develop thoroughly coöperative and effective work together. An author index, pp. 451-456, and a well made subject index, pp. 457-486, complete the book.

A. S. EDWARDS

The University of Georgia

J. E. WALLIS WALLIN. *Education of Mentally Handicapped Children*. New York: Harper and Brothers, 1955, pp. 485.

At a time when many communities are beginning to develop their classes for mentally handicapped children, and when many are attempting to do this work with insufficient comprehension of the many facts needed to do it successfully, this is a most timely book. It has been done by the one man who is perhaps best prepared to do it. It has been done thoroughly well.

Limited to the field of the mentally handicapped, it deals historically with the introduction of this educational work: education adjusted to individual needs; work in the public school special classes; organization and administration of the special classes (two chapters); special class efficiency and teaching procedures; qualifications of the special class teacher; objectives and curriculums for the special class (three chapters); attempts to do this work in regular classes; implications and consequences of mental deficiency; and ultimate aims in the field of mental deficiency.

Wallin assumes that "... the primary agency for the education

of all kinds of children is the publicly supported day schools; and the primary agency for the care, custody, and support of every child is his own (or foster) home. Accordingly, no child should be committed to a residential institution if he can be successfully cared for and educated at large in society. Therefore, the primary agency for the education of mentally deficient or mentally handicapped children is the public school system; state residential schools should serve merely as secondary or ancillary lines of attack or defense." (p. 67.)

But it should be done with proper understanding, efficient organization and administration, and with properly selected and prepared teachers. These matters the author proceeds to discuss with adequate detail and illustration. Methods that have failed and those which have succeeded are both made clear. The reader cannot fail to come to an understanding of why many efforts for the education of mentally handicapped children have failed; and, after a careful reading, cannot lay down the book without a reasonably good insight into certain conditions necessary for success in this field of education. Superintendents and principals will be specially interested in the discussions of organization and administration of these classes. Teachers will be especially interested in the discussions of class room procedures. Students of education will be interested in a competent statement concerning a fast-developing field which they may wish to enter as special teachers. Many parents should be interested in a sound statement of educational principles that may have a far-reaching effect upon some of their children. All who are interested in educational problems will find much that will help them to better understand some of the pressing problems of present day schools.

Many references are given to journal articles and to the best books at the bottoms of pages and at the ends of chapters. The index of subjects includes pages 457-464; the index of names, pages 475-485.

This is not just another book in education. It is a very valuable contribution to a special field of education, either as a text or a reference book. If Boards of Education desire to follow the most promising and economical procedures for success in this urgent developing field, they will do well to take advantage of Wallin's

fifty years of experience as revealed in this volume. The author and publisher are to be congratulated on its publication.

A. S. EDWARDS

The University of Georgia

CLYDE B. VEDDER. *The Juvenile Offender*. New York: Doubleday & Co., 1955, pp. 510.

STARKE R. HATHAWAY and ELIO D. MONACHESI. *Analyzing and Predicting Juvenile Delinquency With The MMPI*. Minnesota: University of Minnesota, 1953, pp. 153. \$3.50.

Delinquency in 1955 is almost everybody's problem. Senators, ministers, newspapers, all express interest in the problem—some in an informed manner and others in a "preachy" moralistic manner. With many people interested, the need for available information to give understanding of the basic problem of delinquent offenders without the need for searching through scientific journals is almost a necessity.

The book edited by Vedder and intended as a supplement to a text book in the field can also serve as a volume containing source material for intelligent lay people interested in the problem of delinquency. The volume is composed of seventy-one articles by fifty-eight contributors. Three contributors rate three articles, two rate two and fifty-three rate one each. The article has introductory comments which set the stage for each of the thirteen chapters. Considered are such topics as: The Extent and Nature of Delinquency; Economic Conditions and Family Factors; Community Institutions; Personality and Behavior Problems; Juvenile Gangs; Juvenile Court; Probation; Correctional Institutions and Parole; and Community Responsibility.

From the Graduate School of the University of Minnesota comes a contribution to delinquency that will be of interest only to students interested in getting at the meaning of the problem by way of measurement. The instrument used here is the MMPI. The authors, to some extent, realize that the instrument used is not in some respects an ideal one for use at Junior and Senior High School levels or, they might have added, the best measuring rod for the study of delinquency at all. However, the instrument is available. Some statistical facts concerning it are known. The use of this instrument under the present circumstances at Minnesota is understandable, as well as perhaps justifiable. In the light of all the clin-

ical and social knowledge that has been coming in, it is doubtful whether this kind of an instrument will be the desirable one to use for the general study and treatment of delinquency. But if by the use of it some basic knowledge can be discovered, studies with the use of this instrument can make a contribution to knowledge in the field.

Seven studies are included in this volume. Some of them have already appeared in journals of psychology or clinical law. The studies included here are: Personality Patterns of Adolescent Girls; Personality Patterns of Juvenile Delinquents as Indicated by the MMPI; MMPI Findings in the Rehabilitation of Delinquent Girls; the Relationship Between MMPI Profiles and Later Careers of Juvenile Delinquent Girls; and Personality Characteristics of Adolescents as Related to Their Later Careers.

One of the outstanding findings of these studies, as interpreted by the editors is the belief that these studies demonstrate beyond reasonable question that some of the patterns of symptoms found in adult mentally ill patients are more common among juvenile delinquents than chance would suggest. That is to say that often delinquents are sick rather than just bad and the implication, of course, is that in professional guidance rather than punishment lies the solution of the problem. For delinquents not classifiable as delinquency-prone with deviant personality, no recommendations appear from these findings. A contention of the editors is that the objective personality test approach is, with all of its limitations, superior to present practices in evaluating need and effectiveness of treatment. This conclusion can be questioned from the point of view of situations as they develop and people as they can be helped. However, it is safe to say that all students interested in the objective approach will find in this volume useful information.

These two volumes about delinquency can serve completely different purposes. The relatively popularized selective readings included in the volume edited by Vedder can serve as reading source material to serve either college students, particularly in colleges where scientific journals are not available, and also intelligent lay people interested in the problem. The Minnesota study, on the other hand, can serve only those people who are interested in the objective approach, the measurement approach, in the study of the problem.

St. Louis, Missouri

H. MELTZER

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

February, 1956

Number 2

EFFECTS OF COÖPERATION AND COMPE- TITION ON THE COHESIVENESS OF SMALL FACE-TO-FACE GROUPS¹

BEEMAN N. PHILLIPS

Indiana State Department of Education

and

LOUIS A. D'AMICO

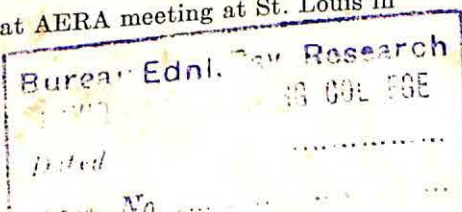
Xavier University, New Orleans

In most classrooms children are encouraged to work both coöperatively and competitively. They are encouraged to get along with others and are even given grades on their report card for "coöperativeness," "helpfulness," "concern for others," etc. But they are also encouraged to compete with others. Goals and standards are set that only a few can reach, and grades and promotions are given on a competitive basis.

This mixture of coöperation and competition has caused much concern among educators in recent years, and an array of conflicting data has been collected with regard to the effects of coöperation and competition on the performance of classroom groups. French (4), Deutsch (2), Stendler, Damrin, and Haines (10) found that classroom groups performed better under coöperative conditions, and Hurlock (6), Leuba (7), Tseng (11), and Whittemore (12) found that classroom groups performed better under competitive conditions.

One of the questions which has been raised concerns the effect of coöperation and competition on group cohesiveness. Research by Mizuhara and Tamai (8), Grossack (5), and Deutsch (1, 2) suggests that coöperation and competition are related to group cohesiveness, but the nature of this relationship is not entirely clear.

¹ Based in part on a paper presented at AERA meeting at St. Louis in 1955.



One of the implications of their findings is that the cohesiveness of a group should increase under coöperative conditions and decrease under competitive conditions. The effect of coöperation and competition on group cohesiveness was the subject of this study, and the hypothesis which was tested was as follows: the cohesiveness of a small face-to-face group will increase under coöperative conditions and decrease under competitive conditions.

PROCEDURE

The subjects used in this study were fourth grade children selected from two schools in a midwest city of 30,000 people. Eight groups were set up with five children in each group. There were four high cohesive groups and four low cohesive groups. These groups were randomly assigned to work under either coöperative or competitive conditions. Two high cohesive groups worked under coöperative conditions and two worked under competitive conditions. Similarly, two low cohesive groups worked under coöperative conditions and two worked under competitive conditions.

According to Festinger, Schachter, and Back (3) the cohesiveness of a group may be based on personal attraction, prestige, and/or goal mediation. In this study group cohesiveness was based only on personal attraction, and high and low cohesiveness were produced by manipulating the degree to which members of a group were attracted to each other. Attraction was measured by a sociometric questionnaire in which each child was asked to choose three children in the room that he would like to sit by. The questionnaire was administered by the teacher at the beginning of the experiment and essentially the same questionnaire was administered at the end of the experiment.

The groups were formed on the basis of the choices made on the initial questionnaire. High cohesive groups were formed by putting together individuals who had selected each other on the sociometric test, while low cohesive groups were formed by putting together individuals who had not selected each other. Thus, it was possible for a group to have from 0 to 15 within-group choices. The actual number of within-group choices for the eight groups on the initial and final sociometric questionnaire is shown in Table 1.

A change in cohesiveness was defined as a change in the number of within-group choices between the initial and the final question-

TABLE 1—NUMBER OF WITHIN-GROUP CHOICES ON THE INITIAL AND FINAL SOCIOMETRIC TEST FOR THE EIGHT GROUPS IN THE EXPERIMENT

Group*	Class	Initial	Final	Level of Confidence
HC-CO	A	11	10	—
HC-CO	B	9	11	.01
HC-CM	A	10	7	.01
HC-CM	B	14	14	—
LC-CO	B	0	4	.01
LC-CO	A	0	3	.01
LC-CM	A	3	3	—
LC-CM	B	0	3	.01

* HC = high cohesion; LC = low cohesion; CO = coöperation; CM = competition.

naire. For example, if a group started out with 10 within-group choices and had 15 within-group choices at the end of the experiment this would indicate that the cohesiveness of the group increased during the experiment.

Coöperation and competition were defined in terms of how members shared in their group's rewards. In coöperative groups they shared equally and in competitive groups they shared in accordance with their relative contributions.

The experimental task was a modification of the game "Twenty Questions." The object of the game was to identify animals by asking the experimenter questions that could be answered yes-or-no. The animals were selected from third grade readers used in the schools. Participation was considered as a regular class assignment, and every pupil in each room participated even though he may not have been part of an experimental group.

Each group was taken to the experimental room and was given the opportunity to identify five animals on each of four successive days. Two randomly selected lists of animals were used, and the order of presentation of the animals was varied. Twenty pieces of candy were distributed among members of each group after identification of each animal. In the coöperative groups each member received four pieces of candy each time an animal was identified. In the competitive groups a scale was used to determine the amount that each member received. The scale which was used is shown

below:

<i>Type of question</i>	<i>Points</i>
Question identifying correct animal	5
Question identifying outstanding characteristics of the animal	4
Question identifying distinctive characteristics of the animal	3
Question identifying common characteristics of the animal	2
Question identifying an incorrect animal	1

Each question asked by a member of a competitive group was assigned a point value. These points were totaled and averaged for each member after an animal was identified correctly. The 20 pieces of candy were distributed in the following pattern: 6-5-4-3-2. The member with the highest number of points per question got 6 pieces, the one with the next highest number got 5 pieces, and so forth.

RESULTS

The results of the study are summarized in Table 1. The number of within-group choices on the initial and final sociometric test are shown for all groups. The symbols HC and LC indicate high and low cohesiveness, and the symbols CO and CM indicate coöperation and competition. Class A had 29 pupils and class B had 38 pupils. In determining whether or not the changes in the number of within-group choices were statistically significant at the one per cent level, only approximate probabilities were computed.

Two significant findings are revealed in Table 1. First, groups which worked under coöperative conditions during the experiment increased in cohesiveness. In three of the four groups the increase in the number of within-group choices was significant well beyond the one per cent level of confidence. This means that individuals who worked together under coöperative conditions liked each other better at the end of the experiment than they did at the beginning of the experiment.

The second significant finding is that groups which worked under competitive conditions did not necessarily decrease in cohesiveness. In two groups there was no change in the number of within-group choices, in one group there was a significant increase in the number of within-group choices, and in the other group there was

a significant decrease in the number of within-group choices. The inconclusive nature of these results suggests that whether or not competition decreases a group's cohesiveness depends on factors not specifically controlled in this investigation.

An idea of the nature of such factors was obtained by a further examination of the data for competitive groups which revealed that rewards were fairly evenly distributed in the groups in which there was either no change or an increase in cohesiveness during the experiment. This suggested the hypothesis that the effect of competition on group cohesiveness is dependent on the effect that competition has on the distribution of the group's rewards. If competition results in more or less uniform distribution of the group's rewards the effect of competition on the group's cohesiveness may be similar to the effect that coöperation would have on the group's cohesiveness. But on the other hand, if one or two members receive most of the group's rewards the effect of competition may be to decrease the group's cohesiveness.

One of the implications of these findings is that classroom groups which are operated on a coöperative basis can be used to improve interpersonal relationships. Such a procedure appears to be potentially the most useful when boys and girls do not know each other very well, when it is desirable to break up cliques or transfer friendships, or when it is necessary to help a new pupil or a shy pupil to establish friendships.

Another implication of these findings is that competition does not necessarily have undesirable effects on interpersonal relationships. The effect of competition on a classroom group's interpersonal relationships seems to be conditioned by its effect on the distribution of the group's rewards. If the members are well-matched and rewards are evenly divided among them, competition appears to have fewer undesirable effects on interpersonal relationships than if members are poorly-matched and rewards are as a result not evenly divided among them. In view of this it may be important in setting up classroom groups on a competitive basis to put individuals together who are more or less equal in ability on the task that has been assigned. In this way it is possible to utilize the incentive of competition in group work without seriously affecting member relationships.

The possible effects of a different type of task, an older group, and a longer period of time in the experimental situation, need to

One group of *Ss* was introduced to the physics material in the "modern" problem-oriented manner. They were first shown a torsion pendulum which consisted of a bar suspended in a horizontal plane from a string. There were four hooks attached to the bar from which weights may be suspended. Two hooks were symmetrically located near the middle of the bar and two were symmetrically located near the ends of the bar. Repeated rotation of the bar would produce a torsion tension in the string which would spin the bar when it was released. The problem for the *Ss* of the "modern" group was to predict whether the bar would spin more rapidly when weights are attached near the middle or at the ends of the bar. They were then told that an answer to the problem could be derived from the physics text which they were to study. They were also instructed that a retention test covering the physics material would be given at a later time.

A second group of *Ss* was presented the study material in the "traditional" manner. They were simply instructed to study the physics text and that they would receive a retention test at a later time. The torsion pendulum was presented near the conclusion of the study period to illustrate a point in the text. Several days later both groups received the two-sphere problem. Significantly more *Ss* in the "modern" method group than in the "traditional" method group solved the problem.

One purpose of this study was to repeat Szekely's experiment with additional control groups in order to obtain further information concerning the variables influencing performance on the two-sphere problem. A second purpose of this study was to determine the relationships between a test of abstract reasoning, certain personality scales, and performance on the two-sphere problem.

METHOD

Subjects. A total of eighty volunteer *Ss* from introductory psychology classes served in the experiment. There were thirty-five men and forty-five women distributed approximately proportionately among the different conditions. The three experimental groups and one control group each contained twenty *Ss*.

Procedure. Approximately two to twelve weeks before the *Ss* served in the experiment, the Taylor Manifest Anxiety Scale (7) was administered to their entire class during the regular class hour.

In the first experimental session the Crown Word Connexion

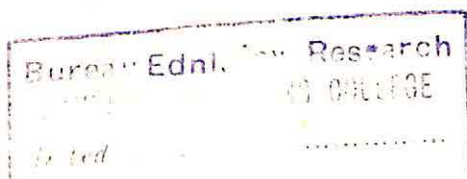
list (2) was initially presented to all Ss. This is a paper and pencil test of neuroticism based on the principle of controlled association. It is composed of fifty items each consisting of a stimulus word and two possible response words. One is considered "normal" and the other "abnormal." Instructions are to underline the response word that has a stronger connection with the stimulus word.

Following this test the Abstract Reasoning Test Form A of the Differential Aptitudes test battery (1) was administered. This is a paper and pencil test in which each of the fifty items consists of a series of four figures, each one of which bears a logical progressive relationship to the preceding one. These are followed by five alternative choice figures. The problem is to choose the figure which continues the progression. Following the abstract reasoning test, one of several different problems was presented to the Ss. They were the x-ray and traffic problems used by Duncker (3) and the water-jar problem of Luchins (4). Szekely (6) also administered a series of problems to his Ss in order to conceal the purpose of the experiment.

For the experimental groups the remainder of this session was devoted to the study of the physics text. The control group received the two-sphere problem at this time. Concepts discussed in the four and a quarter-page text given to the three experimental groups were mass, weight, inertia, acceleration, force, rotational motion, and moment of inertia.³ Szekely (6) states that these are the topics covered in the text of approximately the same length presented to his Ss. Ss were told that they would be tested on the material at the next session and that they could study the text until they felt satisfied that they understood its contents.

The second session for the experimental groups began one or two days later with the presentation of one of Duncker's problems. This was followed by the presentation of the two-sphere problem. Following the presentation of the two-sphere problem the experimental Ss were given a short examination of six questions covering the material in the physics text. The various tests and the test order were the same for the three experimental groups. The only difference among the groups was the manner in which the physics text was presented. All of the Ss served individually in the experiment.

³ We are indebted to Mr. Joel M. Kibbee, teaching assistant in the Department of Physics, for aid in the preparation of the text.



Group I: (Szekely's traditional method). The *Ss* in the first experimental group were asked to read the physics text and were told that they would be tested on the material during the second session. After reading the material they were given an illustration of the moment of inertia using the torsion pendulum. The *E* demonstrated that the pendulum would spin more rapidly when two weights are suspended near the center of rotation than when they are suspended near the periphery.

Group II: (Szekely's modern method). The *Ss* of the second experimental group were first presented with the torsion pendulum as a problem and asked to speculate as to whether it would spin more rapidly with the weights suspended near the center or the periphery. They were then given the text and told that a precise explanation of the principle would be found there.

Group III: The third group received a combination of the two previous procedures. They were first given the torsion pendulum problem as in Group II, then the text, and then the torsion pendulum again as an illustration of the moment of inertia. This variation in procedure was employed in order to more adequately determine the effects of the torsion pendulum illustration on subsequent problem-solving behavior.

Although the solution to the two-sphere problem was dependent upon the material previously studied, this fact was not brought to the attention of the experimental *Ss*. It was given as just another problem in the series of problems presented. The *S* was first asked to visualize two spheres of equal weight and equal size, one of a heavy metal and the other of a light metal. He was asked how this condition could exist. At the same time the *E* drew circles of equal diameter to represent the two spheres. After the solution was verbalized, the *E* differentially shaded the two circles to indicate that one was hollow and that one was solid. The *S* was then told that he would be given thirty minutes to determine a method for differentiating the two spheres without using any special apparatus or damaging the spheres in any way. He was told that there was a way of doing this, for example, right at the table at which he was seated. During the thirty-minute period the *E* answered all questions and indicated improper or unacceptable solutions. A record of the *S*'s responses and questions was kept as well as the time at which they occurred.

RESULTS AND DISCUSSION

Simple analyses of variance of the various scale scores obtained by the different problem-solving groups gave an F of 2.14 for anxiety, 0.53 for neuroticism and 1.52 for abstract reasoning. None of these F values are significant at the five per cent level of confidence, indicating that the variables presumably measured by these tests were randomly distributed among the different problem solving groups. The retention test based on the physics material gave mean scores of 2.9, 2.8, and 2.8 for groups I, II, and III respectively, indicating no difference in retention among the three experimental groups.

The performance criterion employed in the analysis of the results of the two-sphere problem was simply success or failure. These categories were distributed among four classes of responses. S may never suggest using an inclined plane, and would be credited with a failure at the end of the thirty-minute test period. Such S s spend the entire period giving suggestions such as tapping the spheres, bouncing them on the floor, heating or scratching them, and so on. If S suggested rolling the spheres down an inclined plane, three possible response classes may be elicited. He may state that the hollow sphere would roll more slowly, or the solid more rapidly. These were the only responses considered correct. Additional responses to the inclined plane situation that are incorrect are that the spheres would roll at the same rate, and that the hollow sphere would roll more rapidly or the solid sphere more slowly.

Twenty per cent of the control group solved the problem correctly while the percentage of S s solving the two-sphere problem in each of the three experimental groups was fifty, fifty, and forty, respectively. The chi-square obtained from the analysis of these three groups was not significant.

In order to determine whether studying the physics text facilitated solution of the two-sphere problem, the frequency of successes and failures in the combined experimental groups was compared with that of the control group. A significant chi-square was obtained, 5.62, $P = 0.02$.

The results obtained from the experimental groups are obviously contrary to those of Szekely. Twenty per cent of the S s in his group I (traditional method) solved the problem while sixty-five

per cent of the Ss in his Group II (modern method) solved the two-sphere problem. The failure to obtain a difference between these two Groups in the present experiment as compared with Szekely's seems due largely to the superior performance of our group I. The relatively poor performance of group I in his experiment is apparent in that its frequency of success is identical with that of the control group in the present study. Since this group did not receive the study material, their performance presumably reflects the frequency of chance solutions. Regardless of the various hypotheses that may be offered to account for the discrepancy in the results of the two experiments, we may conclude that Szekely's findings are of limited generality. Further speculation would be pointless until additional research can be conducted designed to further investigate the effects of different methods of learning on subsequent problem solving.

Forty-one per cent of the Ss had one or more courses in high school or college physics. They were randomly distributed among the four conditions. Forty-six per cent of the experimental Ss with course work in physics solved the problem. Forty-five per cent of the experimental Ss without courses in physics solved the problem. A chi-square of 0.04 was obtained for this difference and is not significant. However the experimental Ss with prior courses in physics had a mean retention score of 3.38, S.D. = 1.52, while the Ss without prior work in physics had a retention score of 2.41, S.D. = 1.37. The t obtained for this difference was 3.09, $df = 58$, $P = 0.01$. These results show that prior training in physics produced a significant increase in retention but did not increase the number of solutions to the two-sphere problem.

Turning now to the relationships between performance on the two-sphere problem and the other tests, it was found that Ss of the experimental groups scoring above the median, 37, on the abstract reasoning test produced significantly more solutions than Ss falling below the median. Chi-square equaled 5.57, $P = 0.02$. Analyzing the relationship in another manner gives essentially the same results. The mean abstract reasoning score of the twenty-eight solvers was 37.18, S.D. = 6.64; that of the non-solvers, 32.56, S.D. = 9.58. The t for this difference is 2.16, $df = 58$, $P = 0.05$. No relationship between abstract reasoning scores and retention of the physics text was apparent.

A product moment correlation and a correlation ratio were com-

puted between the scores on the neuroticism scale and the abstract reasoning test following their conversion into McCall T-scores. A product moment correlation of 0.01 was obtained which obviously does not differ significantly from zero. But the correlation ratio was 0.67 which is significant, $P = < 0.01$. High scores on the abstract reasoning test tended to be associated with medium neuroticism scores. Low abstract reasoning scores tended to be associated with high and low neuroticism scores. However the chi square analysis of the relative frequency of success on the two-sphere problem and low, medium, and high levels of neuroticism was not significant. Chi-square equaled 0.24.

Analysis of the effects of anxiety indicated that it was not significantly related to any of the other measures studied. The product moment correlation and the correlation ratio between anxiety and abstract reasoning scores following their conversion to T-scores were 0.07 and 0.38, respectively. Neither of these differed significantly from zero. There was a tendency for medium anxiety to be associated with high abstract reasoning scores. The percentages of success on the two-sphere problem at low, medium, and high levels of anxiety were twenty-six, fifty, and fifty-four, respectively. The low anxiety category ranged from one to ten, medium from eleven to twenty, and high anxiety ranged from a score of twenty-one and up. The chi square obtained here equaled 3.26 which is not statistically significant. But again there is a tendency for low anxiety to be associated with failure in problem solving.

SUMMARY AND CONCLUSIONS

Szekely's (6) experimental evidence on problem solving on the two-sphere problem following different methods of learning is contrary to Maier's (5) theory of productive thinking. However, these results were not reproducible in the present study. None of the three different methods of presenting the physics material influenced subsequent problem solving or retention.

It was found, however, that success on the two-sphere problem was related to superior retention. Previous course work in high school or college physics was related to significantly greater retention of the physics material, but it did not influence success on the two-sphere problem. Ss successfully solving the two sphere problem had significantly higher abstract reasoning scores than the Ss that failed.

Crown's (2) neuroticism scale was related in a curvilinear manner to performance on the abstract reasoning scale but was unrelated to performance on the two-sphere problem and the anxiety scale. This absence of any relationship between the neuroticism and anxiety scales is somewhat surprising in view of the common clinical assumption that anxiety is a basic component of neurotic behavior (8). Performance on the manifest anxiety scale was not significantly related to performance on either the two-sphere problem or the abstract reasoning test.

REFERENCES

- 1) G. K. Bennett, H. G. Seashore and A. G. Wesman, *Differential Aptitude Tests*. New York, Psychological Corp., 1947.
- 2) S. Crown, "The Word Connexion List as a Diagnostic Test," *British Journal of Psychology*, 43: 103-112, 1952.
- 3) K. Duncker, "On Problem Solving," *Psychological Monographs*, 58: No. 5 (Whole No. 270), 1945.
- 4) A. Luchins, "Mechanization in Problem Solving," *Psychological Monographs*, 54: No. 6 (Whole No. 248), 1942.
- 5) N. R. F. Maier, "The Behavior Mechanisms Concerned with Problem Solving," *Psychological Review*, 47: 43-58, 1940.
- 6) L. Szekely, "Productive Processes in Learning and Thinking," *Acta Psychologica*, 7: 388-407, 1950.
- 7) J. A. Taylor, "A Personality Scale of Manifest Anxiety," *Journal of Abnormal and Social Psychology*, 48: 285-290, 1953.
- 8) R. W. White, *The Abnormal Personality*. New York, Ronald, 1948.

STUDENT ACHIEVEMENT AS A MEASURE OF INSTRUCTOR EFFECTIVENESS¹

JOSEPH E. MORSH, GEORGE G. BURGESS and PAUL N. SMITH

Personnel Research Laboratory, Air Force Personnel and
Training Research Center

The identification of the successful instructor presents problems which have long been recognized. Over the past fifty years several hundred instructor evaluation studies have been made. The criterion of instructor effectiveness most widely used in these investigations has been some form of rating. Usually supervisor ratings have been used; but peer ratings, student ratings and self-ratings have also been tried out as criteria.

While the studies appear to show that instructors can be reliably rated, ratings are not completely satisfactory as criteria of instructor effectiveness. Ratings are subjective, can be biased, and may not be based on the more important aspects of instructor performance. Dissatisfaction with supervisor ratings, consequently, has led to the search for other criteria that would in some measure at least, avoid the deficiencies inherent in ratings. It is generally agreed that student growth, or gains (the amount of subject matter the student learns with respect to his ability) is one of the most, if not the most, important criterion of instructor effectiveness. There is also general agreement that it is one of the most difficult criteria to measure because of the many controls required to make comparisons meaningful.

Any determination of gain is dependent upon the availability of valid and reliable instruments for measuring student achievement. Proper precautions must be taken to eliminate the possibility of contamination of test results or of the rating measure used. The supervisor who rates the teacher should not also evaluate the students' performance. Tests should have ample ceiling so that all students can demonstrate gain.

Most investigators of student gain as a measure of instructor

¹ This investigation was carried out under the Air Force Personnel and Training Research Center in support of Project No. 7950. Permission is granted for reproduction, translation, publication, and use or disposal in whole or in part by or for the United States Government.

effectiveness have been confronted with sampling problems. In general, student gain in one subject matter area cannot be compared with gains of different students in other areas, and often it has not been possible to secure adequate samples of teachers all imparting the same subject matter to the same kind of students in the same kind of school situation. Consequently, in some studies the numbers of teachers and of pupils have been so small that any findings must be regarded as highly tentative.

If student gains are to be used as a measure of instructor effectiveness, it is necessary to hold constant, insofar as possible, all relevant variables other than the effects of the teaching itself. The instructor is only one of many factors operating to produce changes in the students. Certain aspects of student change such as influence of previous instructors and personal and environmental factors may outweigh the effects of the instructor's performance. Then too, the instructor is called upon to accomplish many changes in his students which are not measurable in terms of subject-matter achievement. Any measure of student gains, therefore, reflects only partially the instructor's total effectiveness.

If it can be determined that students of one instructor make greater gains than do those of another instructor, we can attack the problem of determining what behaviors, traits or characteristics of the successful instructor are responsible for the changes produced in the students.

DESIGN OF THE RESEARCH

In planning this study a course was required in which an adequate sample of instructors could be compared on the basis of their students' achievement. From the available courses in the Air Force technical schools, the Aircraft Mechanic course at Sheppard Air Force Base was selected, and from this course the Hydraulics Phase, taught by one hundred and twenty-one instructors, was chosen as best fitting the requirements of the study. This phase was taught in a single building in which two rows of classrooms were arranged on either side of a central corridor. On each side, the first classroom had training equipment for the first day's training, the second classroom had equipment for the second day's training and so on for the seven days of training in the phase. Every week day six new classes of about fourteen students each

entered the first day of training, two classes on each of three shifts. Each class was assigned an instructor who moved along with his class through the seven days of training. The eighth day was devoted to written and performance testing over the subject matter of the previous seven days. The instructor did not participate in the testing.

Because the Hydraulics Phase lasted only eight days (including the testing day), it was feasible to duplicate the experiment, thus obtaining data based on two classes for each instructor. The correlation between the gains of the instructors' first class and of their second class (taught approximately a month later) was used in determining the reliability of this measure of instructor effectiveness.

STUDENT VARIABLES

Since matching of students to insure classes of approximately equal ability presented almost insurmountable difficulties, it was decided to make adjustments statistically rather than experimentally. For this purpose a special sixty-five item multiple choice pre-test of student knowledge about hydraulics was developed and administered to all students before they entered the Hydraulics Phase. Since the period of instruction in hydraulics was only eight days, the use of identical or even similar tests might have introduced considerable memorization and practice effects. The pre-test, therefore, was composed of items appropriate for students who had had some experience with the subject matter to be taught but had not been exposed to the specific course subject matter. It contained items pertaining to the background knowledge and theory required for learning the phase content as contrasted with the post-test which consisted of items covering specific subject matter information expected of students who had just completed the phase. The content of the pre-test was selected to correlate maximally with the post-test but was not equivalent to it. After statistical determination of internal consistency and item validity as determined by correlation with the post-test scores of a sample of students not used in the experiment proper, four forms were constructed. The four forms contained identical items but with varied order of items or alternatives to reduce the possibility of compromise of the key. That compromise was successfully elim-

inated was shown by the fact that mean scores of subsequent classes did not increase beyond chance expectancy throughout the course of the experiment.

The written hydraulics post-test was composed of seventy-five multiple choice items which satisfied the conditions of the experiment with respect to difficulty, adequacy of incorrect alternatives, and internal consistency. Four forms were constructed in a manner similar to that of the pre-test. This hydraulics post-test was administered to all students (over three thousand) in the experimental group as their final written test in the Hydraulics Phase. All written pre-tests and post-tests were administered and scored by research personnel and were not seen by any hydraulics instructor.

The hydraulics performance examination, a job-sample type test, regularly used in the Hydraulics Phase was included in this study. It concerned checking units for internal leakage, identifying units and fittings, and making operational checks of various hydraulics systems. Each student was required to complete ten of the twenty-eight performance items available. The time varied from ten to thirty minutes per item, the total practical test time being about four hours. Hydraulics instructors trained in the administration of the performance test were used as examiners. In no case, however, did the same man participate in the study as both instructor and examiner. Each item was graded from one to five points. In determining final performance grades, total raw scores were converted to *T*-scores.

The other student variables were the three final phase grades, one each for fundamentals, structures, and electrical, the three previous phases of instruction in the Aircraft Mechanics course.

INSTRUCTOR VARIABLES

Supervisors' ratings of the hydraulics instructors were obtained by means of a form officially designated for use by the Air Training Command. It included both a forced choice and a graphic rating. Shift supervisors were also asked to rank-order the approximately forty instructors on each of the three shifts in terms of their "general effectiveness." Each instructor was assigned a *T*-score based on rank and group size.

Hydraulics instructors were asked to rank-order their colleagues on the same shift in terms of their effectiveness as instructors,

omitting those whom they did not know well enough to rank. The rankings were done anonymously. Mean ranks were obtained and converted to *T*-scores.

Just before taking their final phase examination, students made an over-all rating of their instructors. They also rated them as outstanding, very good, good, poor or unsatisfactory on: (1) knowledge of subject, (2) teaching methods, (3) understanding of students, (4) as a personal friend. They then ranked these four qualities in the order of their instructor's relative strength in them. The instructors were scored separately on each of the four qualities by averaging over students of the two classes to produce two mean scores on each quality, one for the rating and one for the ranking.

The Wonderlic Personnel Test was used as a measure of instructor general intelligence.

Instructor subject matter knowledge was measured in terms of a one hundred and twenty-five-item multiple choice proficiency examination developed for airplane hydraulic specialists.

Verbal facility ratings were obtained by having six supervisors rank-order, in groups of six, individual instructors on their organization and presentation of special material which they had been allowed fifteen to twenty minutes to prepare. Verbal facility scores were determined by averaging the ranks assigned by these supervisors and converting to *T*-scores.

These instructor variables were evaluated in terms of three instructor effectiveness scores. These latter scores were based upon the average student post-test scores adjusted for differences in the average pre-test and ability scores. The adjustment was based on a multiple regression equation computed from within instructor and between classes variances and covariances. Two effectiveness scores were computed for each instructor, one based on the written post-test and one based on the performance post-test. Since these scores had variances of similar magnitude and lacked a basis for arbitrary weighting, a combined effectiveness score was obtained by weighting them equally.

RESULTS

The reliabilities of instructor effectiveness scores were estimated from the correlation between scores based upon the achievement of the first class and those based upon the achievement of the second class.

In terms of the one hundred and six classes in each group, first class written gains correlated with the second class written gains $r = 0.34$; first class performance gains correlated with second class performance gains $r = 0.32$ and when the two measures were combined the coefficient was 0.38. When adjusted for double length by means of the Spearman-Brown formula, these coefficients became 0.51, 0.48, and 0.55 respectively. These latter coefficients represent the estimate of reliability when the two classes for each instructor are used to obtain an effectiveness score.

The correlation coefficients reported below are based on the one hundred and six instructors for whom complete data were available. With this sample size a coefficient of 0.25 is significantly different from zero at the 0.01 level of confidence while 0.19 is significant at the 0.05 level. The written and performance post-test gains were correlated 0.55.

Student Ratings. Students' overall rating of instructors was correlated significantly with all three of the gains criteria (written $r = 0.32$, performance $r = 0.39$, and combined gains $r = 0.40$). Slightly higher correlations were found for students' ratings of instructors' teaching ability, the r 's being 0.41, 0.41, and 0.46, respectively, for the rating and 0.41, 0.42, and 0.47 for trait ranking with the three gains criteria.

Students' ratings of instructors' understanding of students were correlated significantly with the three gains measures (r 's = 0.24, 0.26, and 0.28) but students' rankings of this trait failed to correlate significantly with any gains criterion.

Students' ratings of instructors' knowledge of subject were correlated significantly only with performance gains ($r = 0.22$). Students' ranking of this trait showed a significant negative correlation with written gains ($r = -0.20$) and with the combined gains criterion ($r = -0.22$).

Students' rating of the instructor as a friend was not significantly correlated with any of the three gains criteria, while their ranking of the friendship trait was correlated negatively with all three gains measures (r 's = -0.40 , -0.22 , and -0.36 , respectively). The negative correlations found in the case of the trait-rankings probably represent an artifact due to the requirements of the ranking technique. The sum of the correlation coefficients of the four trait-rankings with an outside variable should approximate zero

TABLE I.—RELIABILITY OF STUDENT RATINGS

(N = 106)

Student Rating	r^*	r^\dagger
1. Over-all (fortunate in having this instructor)	0.26	0.41
2. Teaching ability (rating)	0.34	0.51
3. Teaching ability (trait-ranking)	0.32	0.49
4. Understanding of students (rating)	0.26	0.41
5. Understanding of students (trait-ranking)	0.42	0.60
6. Knowledge of subject (rating)	0.33	0.49
7. Knowledge of subject (trait-ranking)	0.32	0.49
8. As a friend (rating)	0.18	0.31
9. As a friend (trait-ranking)	0.21	0.34

* Correlations between mean ratings by first class and second class.

† r adjusted by Spearman-Brown formula for double length.

because a high score on one trait is compensated for by a low score on another trait.

Since the magnitude of the relationship between student ratings of instructor effectiveness and student gains was encouraging, the reliability of student ratings was estimated by computing the correlation between instructor ratings made by the first class and by the second class. These results are shown in Table I.

The high correlation between student gains and student ratings, however, could be accounted for by factors specific to the student or to the class situation with little reliability over time. To test this possibility the correlations between first class student ratings and second class gains and vice versa were determined. First class student gains correlated significantly with second class student ratings of teaching ability $r = 0.34$, with student ranking of teaching ability $r = 0.25$, and with over-all student rating $r = 0.29$. Second class student gains correlated with these same first class measures $r = 0.08$, $r = 0.25$, and $r = 0.09$, respectively. The true values of the correlations possibly lie somewhere between these two sets of values. If so, some real validity can be attributed to the student ratings.

All student rating variables tended to be highly intercorrelated, with the students' trait-ranking of the instructor as a friend and instructor's knowledge of subject matter tending to give signifi-

cantly negative correlations. As explained above, these negative correlations were probably artifacts of the trait-ranking technique.

Peer Rankings and Supervisor Ratings and Rankings. Neither peer rankings nor supervisor ratings nor rankings were correlated significantly with any of the three student gains criteria. The two supervisor measures when correlated yielded a coefficient of 0.67. Supervisor-forced choice ratings were correlated with peer-ranking ($r = 0.52$) while supervisor-ranking and peer-ranking correlated 0.77. This high correlation found between supervisor and fellow instructor rankings coupled with the fact that neither of these measures correlated highly with the student gains criteria suggests that both supervisors and peers judged instructors on the basis of factors other than teaching effectiveness as measured by the gains criteria. As will appear later, one of these factors seems to be the instructors' knowledge of subject matter. We thus see rather close agreement between peer and supervisor opinion but fellow instructors and supervisors agree only slightly with student opinion.

Other Measures. Verbal facility ratings correlated significantly with the written gains measure ($r = 0.20$) but with neither of the other gains criteria. Neither instructors' hydraulics subject matter knowledge nor intelligence were correlated significantly with any of the student gains criteria. We see, therefore, that student ratings of their instructors were the only instructor measures which seemed to predict the student gains criterion.

Verbal facility, however, was correlated positively with scores on the Wonderlic test ($r = 0.23$) but with no other instructor measure. There was a correlation of 0.44 between intelligence test scores and the hydraulics proficiency examination. A correlation coefficient of 0.43 was found between scores on the proficiency test and students' ranking of instructors' knowledge of subject matter, and an r of 0.19 between the proficiency test and students' rating of instructors' knowledge of the subject. There was a negative correlation ($r = -0.22$) between the proficiency test and students' rating of the instructor as a friend, the correlation not being significant in the case of the friendship trait-ranking. Students' trait-ranking of the instructors' understanding of students also correlated negatively ($r = -0.27$) with instructors' proficiency test scores. Here again, these negative correlation coefficients may be due to the trait-ranking technique used.

IMPLICATIONS

There is a continuing need for measures to be used in the evaluation of instructor performance. Such standards of success in teaching are required so that the most effective instructors may be promoted, upgraded, or given supervisory responsibility while the less effective instructor may be given remedial training or reassignment.

The present study was aimed at developing a reliable criterion of instructor effectiveness. Although the work was done with instructors in an aircraft mechanics course at an Air Force installation, it was anticipated that the results of the study would find application in other teaching situations.

Because it provides the most objective criterion so far discovered, the effectiveness measure used was the actual subject-matter achievement of an instructor's students. Since a precise comparison of instructors in terms of student achievement is usually impractical, an attempt was made in the study to find other characteristics of the instructor, besides ability to impart subject matter, which might be related to student gains and could thus be used to predict student achievement.

The chief results of the investigation were the findings that student gains can be reliably measured and that students' ratings of their instructors' teaching effectiveness, and supervisors' rating of instructors' verbal facility are correlated significantly with student gains.

Technical school students appear to know when they are well taught. Their ratings of their instructors offer promise as a technique of instructor evaluation. Teaching which emphasizes acquisition of subject matter may induce a favorable attitude of students toward their instructors.

The high relationship found between ratings and rankings by fellow instructors and supervisors, together with the fact that these measures appear unrelated to student gains, suggests that fellow instructors and supervisors judge instructor effectiveness on the basis of factors other than what students learn. One of these factors appears to be the instructor's knowledge of subject matter. The study also suggests that further, more detailed investigation might be made of speech factors as related to instructor effectiveness.

SUMMARY

(1) This study showed that under some conditions, student gains can be reliably measured.

(2) The students appeared to know when they were well taught. Student ratings, therefore, offer promise as a technique for instructor evaluation.

(3) The students' rating of instructors' subject-matter knowledge was correlated significantly with instructors' proficiency test scores.

(4) Little relationship was found between student gains and instructor intelligence or knowledge of subject matter.

(5) Little relationship was shown between supervisor or fellow instructor estimates of instructor effectiveness and student gains.

(6) The high correlations found between fellow instructor and supervisor rankings plus the fact that neither of these measures correlated highly with student gains suggests that fellow instructors judge instructor effectiveness on the basis of factors other than student achievement. One of these factors appears to be subject-matter knowledge.

(7) The correlation between gains on the written and performance tests indicates that they have about twenty-five per cent common variance. These measures show a logical patterning with other measures which implies that gains scores are specific to the type of post-test used.

(8) The correlation between verbal facility rating of the instructor and the student written gains criterion while not high was significant (at the 0.05 level) and suggests further investigation of speech factors as related to instructor effectiveness.

PERSONALITY FACTORS OF OVER- AND UNDER-ACHIEVERS IN ENGINEERING¹

ELVA BURGESS

Alabama Polytechnic Institute

PROBLEM

The purpose of this study was the investigation of the hypothesis that students who over-achieve in a college situation have common personality factors which differentiate them from college students who under-achieve.

THE SAMPLE

From a population of four hundred and ninety-two male freshmen engineering students for whom predictions of first semester academic performance at the Pennsylvania State College had been made in the fall of 1951, two samples of twenty each were selected for study.

The predictive index which had been used was a multiple coefficient of correlation of 0.62, based on Arithmetic and Algebra subtests of the Moore-Castore Test of Academic Aptitude (9) and high school rank in fifths.

The samples were drawn in the following manner: From the Office of the College Registrar the grade-point average earned by each of the four hundred and ninety-two students during the fall of 1951 was obtained. The mean and standard deviation were computed. Mean grade-point average was 1.10, S. D. 0.93, almost exactly the same as had been previously reported for the engineering freshmen in this College in the fall of 1950.

Applying the range, 1.33 to 0.87, which represents the mean grade-point average plus or minus one-quarter of a standard deviation, to the predictions shown on an alphabetical list of the four hundred and ninety-two students, a second list of one hundred and twenty-eight students was obtained. These represented the one hundred and twenty-eight students whose predicted grades had

¹ This article is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree, The Pennsylvania State College, 1953.

been within one-quarter of a standard deviation from the mean of the actual grades earned.

From the list of one hundred and twenty-eight students, the twenty whose earned grade-point average deviated most above their predicted grades were selected and became one criterion group, called the over-achievers. The twenty students whose actual grades showed the greatest deviation below their predicted grade-point average were selected as the other criterion group, called the under-achievers. The mean grade-point average predicted for the over-achieving group was 1.10, S. D. 0.14; for the under-achieving group the mean was 1.09, S. D. 0.10.

With the exception of one member of the over-achieving group who was twenty-two years of age and a veteran, all students were in their eighteenth or nineteenth year at the time of testing. All were white and unmarried. Each of the individual engineering curriculums was represented in each group.

TEST ADMINISTRATION AND SCORING

A battery of six tests, as follows, was individually administered by the experimenter to all members of both groups, with the exception of one member of the over-achieving group, who completed only the first two tests:

- (1) The Rorschach Technique (11).
- (2) The Minnesota Multiphasic Personality Inventory, Group Form (6).
- (3) The Murray Thematic Apperception Test, Cards 1, 2, 3BM, 8BM, 9BM, 11, 14, 17BM, 18BM, 19, and 16, in that order (10).
- (4) The Rosenzweig Picture Frustration Study (12).
- (5) The Strong Vocational Interest Blank for Men (14).
- (6) The College Inventory of Academic Adjustment (Borrow) (5).

In addition, results of the Bernreuter Personality Inventory (4) for each of the subjects were obtained from the files of the Student Advisory Service of The Pennsylvania State College, and used in the study. This test had been administered by the Student Advisory Service staff a few months previously.

Until all testing, test scoring and checking were completed, the subjects were identified to the examiner only by testing number. This precaution was deemed desirable in order to minimize and equalize as much as possible the effects of the subjectivity which,

by the nature of the instruments, enters into the administration and scoring of the projective techniques, particularly the Rorschach and the Thematic Apperception Test.

The Rorschach Technique was administered in accordance with the method recommended by Beck, his suggested directions having been used as standard directions for this testing (2). Scoring also followed the Beck system, with the exception that the protocols were not scored for his *Z* factor, and were scored for *FM* (animal movement) and *m* (inanimate movement), as is done by Klopfer (8) and Hertz (7). For scoring quality of form, statistical tables published by Beck (2) and by Hertz (7) were consulted, in the order named.

The subjects wrote their T.A.T. stories, with a time limit of five minutes for each, except in the case of Card 16 where some extension of time was granted. The T.A.T. stories were scored for needs, press, and outcome in accordance with the scoring system published by Stein (13).

All other tests were administered in accordance with the directions given in the manual provided by the authors of each test. The Strong Vocational Interest Blank for Men was scored on only two keys, the Occupational Level Scale and the Interest Maturity Scale. All scoring was checked by re-scoring, even the projective techniques.

STATISTICAL ANALYSIS

Means and standard deviations were computed, and a test of the significance of the difference between means of the two groups was made by employment of the *t*-ratio for small samples for the following number of test variables:

- (1) Rorschach Technique: forty-five location, productivity, and determinant variables, plus thirty-five signs.
- (2) The Minnesota Multiphasic Personality Inventory, Group Form: fourteen variables.
- (3) Murray Thematic Apperception Test: seventeen variables.
- (4) The Rosenzweig Picture Frustration Study: twenty-five variables.
- (5) Strong Vocational Interest Blank for Men: two variables.
- (6) Borow College Inventory of Academic Adjustment: seven variables.
- (7) Bernreuter Personality Inventory: six variables.

TABLE I—SUMMARY: TEST FACTORS FOR WHICH A "t" OF 1.0 OR MORE WAS OBTAINED FOR DIFFERENCE BETWEEN MEANS

Test	Over-Achievers			Under-Achievers			t
	N	Mean	S.D.	N	Mean	S.D.	
Rorschach	20			20			
VIII-X per cent		32.10	8.12		38.68	7.54	-2.59
R							
Affective ratio (Beck)		0.49	0.18		0.65	0.19	-2.68
R, total responses		31.15	16.73		37.70	18.60	-1.14
P, popular responses		5.65	2.39		6.60	2.20	-1.28
A% (animal content)		43.00	13.60		51.60	16.45	-1.76
F% (form only)		55.80	16.38		61.25	15.48	-1.05
FC		1.60	0.97		2.35	1.73	-1.65
H + A		13.15	5.77		16.95	8.78	-1.58
Hd + Ad		5.80	5.91		8.30	7.14	-1.18
S% of R		8.40	6.20		5.90	5.30	1.34
V% of shading responses		41.69	32.38		25.56	25.93	1.49
H + A: Hd + Ad	16	2.42	1.17	18	3.91	3.98	-1.48
Fluctuation in productivity		1.34	1.41		1.52	1.62	-1.13
Thematic Apperception Test	19			20			
Ending happy		6.21	2.98		5.25	2.81	1.01
Total needs		19.00	5.43		16.45	4.36	1.57
Total needs and press		23.93	6.94		20.90	6.32	1.40
Needs: hero initiates activities re: situations or objects		10.68	4.05		8.15	3.62	2.00
N achievement		2.16	1.14		0.90	1.09	3.44
Aggressive needs		1.47	1.50		0.55	0.59	2.43
Improvement needs		5.84	2.76		3.25	2.86	2.80
"Freedom" needs		1.26	1.33		1.95	1.66	-1.40
Dependent needs		0.32	0.57		1.30	1.31	-2.97
Rosenzweig Picture Frustration Study	19			20			
I'		0.97	0.64		1.22	0.76	-1.09
M (in E - D column)		3.68	1.62		2.78	1.29	1.86
M (total)		6.63	1.75		5.92	1.67	1.26
M%		27.58	7.07		24.95	7.15	1.12
O - D%		17.42	6.30		19.95	4.64	-1.38
E - D%		54.11	9.74		50.30	6.59	1.38
E%		6.14	3.46		7.22	2.53	-1.08
I%		7.45	4.16		5.43	2.74	1.72
O - D column		4.16	1.50		4.75	1.15	-1.34
E - D column		12.97	2.33		12.02	1.64	1.42
MMPI	20			20			
H _s scale		12.25	1.37		11.05	2.65	1.75
M _f scale		21.70	3.49		23.50	3.74	-1.45

TABLE I—Cont.

Test	Over-Achievers			Under-Achievers			<i>t</i> *
	N	Mean	S.D.	N	Mean	S.D.	
Strong Vocational Interest Blank for Men	19			20			
Occupational level		7.79	47.82		-16.00	53.25	1.43
Borow College Inventory of Academic Adjustment	19			20			
Part I		17.79	4.38		14.70	5.96	1.80
Part II		22.42	4.77		20.05	6.60	1.26
Part III		25.68	5.07		19.30	8.61	2.76
Part IV		29.89	8.29		26.65	9.67	1.10
Part VI		20.95	4.05		18.90	5.37	1.32
Total score		137.26	27.35		118.40	39.02	1.71

* A *t* of 2.72 is significant at the one per cent level; 2.43, at two per cent level; 2.03, at five per cent level; 1.69, at ten per cent level. Negative value indicates that the difference was in favor of under-achievers.

Protocols of the T.A.T. were also studied in an effort to locate differentiating content material.

RESULTS

All factors for which the *t*, obtained for the significance test of the difference between means, was 1.0 or higher are presented in Table I. Statistics for the six Rorschach signs are given in Table II.

For none of the Rorschach *variables* was the difference significant at the one per cent level of confidence. (However, two of the *signs* were significant at this level). Two variables, VIII-X per cent and affective ratio (Beck), were in favor of the under-achievers and significant at a level beyond the two per cent but not reaching the one per cent level of confidence.

Variables for which the obtained *t* was not significant but was greater than 1.0, with the difference in favor of the under-achievers, were: Total *R*, *A* per cent, *F* per cent, *P*, *FC*, *H* + *A*, *Hd* + *Ad*, *H* + *A*: *Hd* + *Ad*, and fluctuation in productivity.

Variables on which the over-achievers scored higher and the *t* was 1.0 or greater but not to a significant level were: *S* per cent and *V* per cent (per cent of vista responses to all shading responses).

The sign-check of the Rorschach protocols revealed six signs which, on the basis of a *t*-test of the difference between percentages, differentiated the two groups. Two of these were negative signs.

TABLE II—SIGNIFICANCE OF RORSCHACH SIGNS

Sign	Over-Achievers (N. 20)			Under-Achievers (N. 20)			t*
	N	Per cent	SE Per cent	N	Per cent	SE Per cent	
<i>Ap</i> is <i>W</i> , (<i>W</i>), or <i>D</i>	14	70	10.25	8	40	10.95	2.00
<i>A</i> %, 50% or less	16	80	8.94	9	45	11.14	2.45
Affective ratio, 0.65 or less	17	85	8.00	10	50	11.18	2.68
No poor form quality (<i>F</i> -) or Pure <i>C</i> on Card IX	17	94†	5.61	5	28†	10.58	5.51
<i>FC</i> equal to or greater than <i>CF</i>	8	40	10.95	14	70	10.25	-2.00
Per cent of responses to last 3 cards, 40 to 60%	3	15	8.00	11	55	11.14	-2.92

* A *t* of 2.03 is significant at the five per cent level; 2.43, at two per cent level; 2.72, at one per cent level. Negative value indicates difference was in favor of under-achievers.

† *N* is 18 because of two card rejections.

The six signs were:

AP is *W*, (*W*) or *D* (Beck), *t* significant just short of the 5 per cent level of confidence.

A per cent, fifty per cent or less, significant at the two per cent level.

Affective ratio, 0.65 or less, significant beyond the two per cent level.

No *F*- or pure *C* on Card IX, significant beyond the one per cent level.

FC equal to or greater than *CF*, a negative sign, significant just short of the five per cent level.

Per cent of responses to last 3 cards, forty to sixty per cent, a negative sign, significant beyond the one per cent level.

When a point-score of 1 was assigned to each and a total sign-score of 4 used as the cutting score for achievement, eighty per cent of the over-achievers were correctly classified and only ten per cent of the under-achievers incorrectly included as over-achievers.

The statistical analysis of the T.A.T. variables revealed that over-achievers scored higher on need for achievement (significant beyond the one per cent level), needs to be aggressive (significant just at the two per cent level), and need to improve self or status (significant beyond the one per cent level). Their needs were more fre-

quently related to activities initiated with regard to objects and situations (significant just short of the five per cent level of confidence). With a t of 1.0 or higher taken as indicator of a tendency, over-achievers showed a tendency to more total needs and more needs and press combined, and to give happy endings to their stories. Under-achievers scored significantly higher (beyond the one per cent level) in number of needs to be dependent, and t -ratios higher than 1.0 but not to the point of statistical significance suggest that they have a tendency to have more need to be free from restraints.

Analysis of Card 2 of the T.A.T. revealed that these male college students readily identified the girl in this picture, eighty-four per cent of the over-achievers and eighty-five per cent of the under-achievers identifying her as a student. Of the under-achievers who told stories about the girl in a school situation, fifty-three per cent attributed her presence in school or college to parental pressure or other sources of motivation, and forty-seven per cent thought she was not interested in school or disliked it. Corresponding percentage for over-achievers was zero in each case, and eighty-three per cent of those telling stories in which the girl was placed in a school situation saw her as positively oriented to school. Forty-five per cent of the total group of under-achievers saw her as preferring to remain in the home environment, and again none of the over-achievers thought that; on the contrary, sixty-three per cent of them ascribed to her a feeling of wanting to escape the limitations of life on the farm.

Card 2 seems an excellent instrument for revealing attitudes toward the school situation, i.e., like or dislike for school, value of a college education recognized or not, and basis of motivation for attendance at college. Attitudes toward environmental factors were also disclosed.

No significant differences were uncovered by the Rosenzweig Picture Frustration Study. Variables on which over-achievers scored higher and the obtained t was 1.0 or greater suggest that this group has a tendency to emphasize the assignment of the blame in frustrating situations but that they are impulsive, i.e., assign blame to circumstances which are unavoidable, and that in super-ego situations they admit guilt but claim extenuating circumstances. Variables on which under-achievers scored higher and for which the obtained t -ratio was 1.0 or more but not to a point of

statistical significance, indicated that this group tends to emphasize the frustrating object or problem in a frustrating situation, and to be more intro-punitive, i.e., to turn the aggression inward, but minimizing the frustration while doing so. In super-ego situations they tend to be more aggressive in their denial of blame when charged with an offense.

No significant differences were obtained for any of the clinical or validating scales of the MMPI or for an Anxiety Index or Internalization Ratio (15). Obtained *t*-ratios of 1.0 or higher suggest a tendency for over-achievers to score higher on the *H_s* scale and under-achievers to score higher on *M_f* scale.

No *t*-ratios as high as 1.0 were obtained on the six scales of the Bernreuter Personality Inventory.

On the Strong Vocational Interest Blank for Men no significant difference was found for either the Occupational Level scale or the Interest Maturity scale. For the OL scale the obtained *t* was 1.43 in favor of over-achievers.

For all seven variables of the Borow College Inventory of Academic Adjustment the direction of the difference between means was in favor of the over-achievers, with *t*-ratios higher than 1.0 in all but one instance, Part V, Mental Health, and significant beyond the one per cent level on Part III, Personal Efficiency in Planning and Use of Time.

CONCLUSIONS

The hypothesis that over-achievers among a group of college students predicted to attain average academic achievement in engineering curriculums would show common personality factors which would characterize them as a group and distinguish them from a group of under-achievers has been substantiated. Specifically, it has been found that:

(1) Over-achievers are less labile in their affective reactions, tend more towards constriction, and are more inhibited in emotional response to pleasurable aspects of the environment. Intellectual adaptivity is greater, the approach to problems is more cautious and concretistic, and intellectual control of emotional reaction in the face of strong outer stimulation is more effective. Need for achievement and improvement of the self or status is greater, and they are more motivated for college study, enjoy it more, and expect to get more from it. They are more efficient in

the planning and use of their time, and tend in general to be better adjusted to the college situation. To a greater degree they chafe under environmental limitations as they have known them and relate attainment of an education to escape from those restrictions. They show more needs to be aggressive, and less social skill.

Expectations that they would show evidence of more feelings of inferiority, a more optimum amount of anxiety, and higher occupational level and maturity of occupational interests were not confirmed. The findings in these respects were not statistically significant, although they did lie in the anticipated direction. The expectation that they would give evidence of more self-sufficiency was not confirmed.

(2) Under-achievers are less intellectually adaptive, over-generalize and over-extend the self, and show less intellectual control and repression of emotional reactivity. They over-react to environmental circumstances, and in general show easy, labile affectivity. Establishment of rapport in social situations is easier, but they are more dependent in their attitudes towards others. Motivation for academic achievement is weak; they tend not to enjoy the school situation, and to be unable to see the value of an education. They tend to see their own environment as a desirable one.

In addition, results obtained in this study seem to justify the following conclusions:

(3) Although the Rorschach Technique, individually administered and subjected to statistical analysis by means of the *t*-test, does reveal certain personality differences between academic over- and under-achieving groups, the testing of the individual variables for significance of difference between means is not adequate to reveal the more subtle personality differences nor the inter-relationships which are sometimes thought to exist between Rorschach variables.

(4) The Thematic Apperception Test, particularly Card 2, is a potentially more useful instrument for a research approach to personality study of academic achievement than is the Rorschach. For intensive and definitive research, however, all of the T.A.T. cards should possibly be administered and subjects should not be asked to write their stories, if the most is to be obtained from this technique.

(5) The Minnesota Multiphasic Personality Inventory, Group Form, the Bernreuter Personality Inventory, the Occupational

Level and Interest Maturity scales of the Strong Vocational Interest Blank for Men, and the Rosenzweig Picture Frustration Study, when given and scored in the conventional manner, are not suitable instruments for differentiating groups of academic deviates.

(6) The Borow College Inventory of Academic Adjustment, particularly Part III, offers sufficient promise as an effective instrument in prediction of college achievement to warrant its wider trial-use in individual counseling and group research for purposes of further exploring its worth.

IMPLICATIONS FOR FURTHER RESEARCH

Results obtained in this study suggest that additional research particularly along the following lines might prove fruitful.

(1) Cross-validation, using larger samples, of the six Rorschach signs. It would appear that the sign, no *F*- or pure *C* on Card IX, showing a high degree of significance in this study, would be especially useful in an individual counseling situation if it were found to hold up under cross-validation.

(2) Another study, also using larger samples, aimed at a more comprehensive investigation of the Thematic Apperception Test as a discriminative technique for academic deviates would be highly desirable. Again, if the clear-cut and specific differences revealed for Card 2 were confirmed in a cross-validation study, this finding would have immediate applicability and utility in individual counseling, particularly when time available with the client is definitely limited by the situation, as it usually is, for example, in Veterans Guidance Centers.

(3) Although the differences found on the Rosenzweig Picture Frustration Study were not significant, the suggestion that over-achievers tend to be more impulsive might be profitably investigated in another study. Impulsiveness has been interpreted as evidence of repression, an interpretation which is generally in accord with the other findings in this study. Research using larger samples and trying out a time limit below the median eighteen minutes taken by the subjects in this study might prove interesting.

(4) Findings for the Occupational Level key of the Strong Vocational Interest Blank for Men suggest that further research might reveal a standard score below 40 on that scale to be indicative of under-achievement. This suggestion, if corroborated by additional

research, would again be highly useful in individual counseling, or perhaps, even in group prediction.

REFERENCES

- 1) S. J. Beck, *Rorschach's Test: II. A Variety of Personality Pictures*. New York, Grune & Stratton, 1947.
- 2) S. J. Beck, *Rorschach's Test: I. Basic Processes*. New York, Grune & Stratton, 1949.
- 3) S. J. Beck, *Rorschach's Test. III. Advances in Interpretation*. New York, Grune & Stratton, 1952.
- 4) R. G. Bernreuter, *Manual for the Personality Inventory*. Stanford, California, Stanford University Press, 1935.
- 5) H. Borow, *Manual for the College Inventory of Academic Adjustment*. Stanford, California, Stanford University Press, 1946.
- 6) S. R. Hathaway and J. C. McKinley, *Minnesota Multiphasic Personality Inventory; Manual*, revised. New York, Psychological Corporation, 1951.
- 7) Marguerite R. Hertz, *Frequency Tables to be used in Scoring Responses to the Rorschach Ink-blot Test*, third edition. Cleveland, Ohio, Western Reserve University, 1946.
- 8) B. Klopfer and D. Kelley, *The Rorschach Technique*. Yonkers-on-Hudson, New York, World Book Company, 1946.
- 9) B. V. Moore, and G. F. Castore, *The Pennsylvania State College Academic Aptitude Examination*, revised. The Pennsylvania State College, State College, Pennsylvania, 1947.
- 10) H. A. Murray, *Manual for the Thematic Apperception Test*. Cambridge, Mass., Harvard University Press, 1943.
- 11) H. Rorschach, *Psychodiagnostics, a Diagnostic Test Based on Perception*, translated by P. Lemkau and B. Kronenberg, fourth edition. New York, Grune & Stratton, Inc., 1949.
- 12) S. Rosenzweig, Edith E. Fleming and Helen J. Clarke, *Revised Scoring Manual for the Picture Frustration Study*. Provincetown, Mass., Journal of Psychology Press, 1947.
- 13) M. I. Stein, *The Thematic Apperception Test. An Introductory Manual for its Clinical Use with Males*. Cambridge, Mass., Addison-Wesley Press, 1948.
- 14) E. K. Strong, Jr., *Manual for Vocational Interest Blank for Men*. Stanford, California, Stanford University Press, January, 1951.
- 15) G. S. Welsh, "An Anxiety Index and an Internalization Ratio for the MMPI," *Journal of Consulting Psychology*, 16: 65-72, 1952.

A RESEARCH CHECKLIST IN EDUCATIONAL PSYCHOLOGY

PERCIVAL M. SYMONDS

Teachers College, Columbia University

For eleven years, from 1944 until the present time, the writer has been chairman of the departmental seminar and project conference in Psychological Foundations in Teachers College, Columbia University. This seminar, representing the most advanced work in the Department of Psychological Foundations, is a requirement for all doctoral candidates; and an acceptable report is one of the requirements for certification for the Doctor of Philosophy degree. It provides an opportunity for students to report on their doctoral problems or projects, and to benefit by discussion, suggestions and criticism from staff and members of the seminar. Parenthetically, it may be noted that at Teachers College the study on which the student works as a dissertation for the Ph.D. degree is called a problem, whereas the field work or plans with regard to an educational enterprise, a report of which the student submits as one of the requirements for the Ed.D. degree, is called a project. The first report that the student makes in this seminar usually consists of his plan for the problem or project. Subsequent reports will register progress on the student's research and will give the seminar the opportunity to think along with the student on the various problems that arise during the course of his investigation.

Students are encouraged to take an active part in the discussions of the seminar, and they are judged in part by their contributions to discussion.

Following the seminar each week the departmental staff meets to discuss the student's presentation, and evaluations are made of the report. Typically the staff appraises the student's plans and makes recommendations for his prosecution of his study.

A report of the discussions in the seminar, prepared by the chairman each week, includes the outline of the study which the student has prepared for the seminar, records all questions raised and the discussion around these questions and in addition the deliberations and conclusions reached by the staff.

The checklist of questions which constitutes the body of this

paper has been derived from these weekly reports. Questions raised during each seminar presentation were taken off on slips and then rearranged by topics. This checklist may serve as an inventory of the kind of issues that arise in planning a research study in contemporary educational psychology and should be of service to students who are planning research projects. If a student answers for himself the questions raised in this checklist, he should be able to satisfy the questions which are likely to be put to him by members of an educational psychology faculty during a seminar report. The checklist ought also to be of some value to those who give instruction in research methods as an outline of the various issues which might be presented in such a course.

QUESTIONS TO BE CONSIDERED IN PLANNING A DOCTORAL
DISSERTATION OR PROJECT IN PSYCHOLOGICAL
FOUNDATIONS

This outline is based on questions, comments and suggestions made in Education 307SC-308SC (Departmental seminar and project conference in educational psychology, developmental psychology, remedial reading in elementary school and psychology of the elementary school subject, social psychology and educational measurement and evaluation, Teachers College, Columbia University, 1944 to 1955.)

The following outline should serve as a checklist of the more common queries raised in the critical evaluation of proposed research studies

A. Scope and Definition of Study

1. Is your problem being considered broadly enough?
2. Have you sufficiently limited your problem?
3. What are the educational implications of your study?
4. Have you governed your decisions by the experiences of investigators who have preceded you?
5. How significant psychologically is a socially selected group such as "children who steal," "children who nail bite," "inmates of a correctional institution?"

B. Hypotheses

1. What are the hypotheses?
2. Are the hypotheses promising?

3. Are the hypotheses clearly and precisely stated?
4. Are the hypotheses stated in a form that permits them to be tested?
5. Is it better to hazard a hypothesis or to ask a question?
6. Has the study been restricted to one or a few principal hypotheses to be tested?
7. Are your hypotheses independent of one another?
8. Is it better to hypothesize causal factors or merely to hypothesize relationships?

C. Background

1. Have you made a thorough, careful review of the literature pertaining to your problem?

D. Definitions

1. Have proper distinctions been made between concepts?
2. Have concepts been adequately analyzed so as to distinguish between small but significant differences in method, materials, subjects, setting, etc.?
3. Is there clear and unequivocal meaning in the use of your terms?
4. Are concepts adequately and accurately defined?
5. Do some of your concepts require restrictive definition?
6. Do the meanings of your terms change with changes in age, sex, socio-economic status, etc.?
7. From what (whose) point of view are you defining your terms?

E. Method of Study

1. Has a decision as to the method of inquiry been made?
2. Is there a relation between the data to be collected and the hypotheses which the study is trying to test?
3. Do you propose to collect your own data or will you make use of published data already gathered?
4. Are you planning to use a "shotgun" approach?
5. If you plan to study individual cases, have you given thought as to how you will go from cases to general conclusions?
6. Are the data necessary for your study available?
7. Are you in a position to secure the data necessary for a successful prosecution of your study?

8. Are there variations in the method whose results would be worth investigating?
9. When more than one investigational approach is available, is it worth while to compare the results using different methods?
10. Can you draw conclusions as to cause and effect from evidence as to relationship?
11. To what extent can you generalize from a single experimental situation?
12. Are you planning to draw general conclusions from a study limited in age, sex, social class, race, etc.?
13. Have you considered the desirability of studying the relationships covering the whole range of your population instead of studying contrasting extreme groups?
14. Is the range among your cases sufficient to permit you to demonstrate the relationships in which you are interested?
15. Is it possible that the differences or relationships in which you are interested might be due to differences inherent in the situation (personality differences, for instance) instead of or as well as the experimental factors you propose to study?
16. How do you propose to select your criterion groups?
17. Can you make detailed observations of individual subjects to supplement your mass data?
18. Would it be well to ask questions on both the positive and negative side of an issue instead of just one side?
19. Are you justified in assuming a constant motivation among your subjects?
20. Will your results be influenced by the order or position of your materials?
21. Have you considered the possibility of response sets in your subjects; i.e., a general tendency to respond, which might influence your results?

F. Design

1. Is the design of your study clearly formulated?
2. Have you taken into account the various hidden factors which might influence the results of your study besides the variables that you are specifically planning to study?
3. Have you taken into account the influence of age, sex, school

- grade, I.Q., socio-economic status, mental set, leakage among subjects, influence of examiner, emotional factors?
4. Have you taken sufficiently into account experiences other than your experimental variables that might intervene between your first and second test?
 5. How may the influence of these variables be eliminated?
 6. Have you made a decision with regard to whether it is better to control variables experimentally or to test the contribution of variables statistically?
 7. Have you given sufficient thought to the necessity of controls?
 8. Are you in a position to measure and/or control variables which might influence your results?
 9. Does the design of your experiment permit you to randomize variables which you do not want to influence your results?

G. Sampling

1. To what extent will you be able to generalize your findings?
2. What are the criteria for selecting your cases?
3. Is sampling of materials, places, subject, test items, etc. adequate?
4. How do you propose to select subjects for your study?
5. Are your proposed subjects representative of the population to which you intend to generalize your results?
6. What age level and age range do you propose to study?
7. Have you given sufficient attention to the sampling of tasks?
8. What factors may be biasing the selection of subjects?
9. Do you propose to use groups already selected by social processes or do you propose to select groups by direct testing of the characteristics of individuals?
10. Would your results be more clear cut if you selected groups which give promise of showing wide differences?
11. How stable will your findings be—that is, will they stand up when made under different conditions—when made with other subjects, materials, instructors, examiners, in other places, etc.
12. Will you make your selections on the basis of known factors or on the basis of a random sampling?
13. Should you treat as a unit individuals over a wide age, grade or cultural range?

14. Can you simplify your problem by limiting it to a narrower age range, grade range, geographical range, etc?
15. Are you taking into account the sub-groups in your total population?
16. How do you propose to group your subjects?
17. In studying the differences between two populations, are you sure the populations are differentiated on the variables which you assume differentiate them?
18. Are you determining your sampling or are you letting others select cases for you?
19. If you depend on voluntary participation for your subjects, have you given consideration to what this will do to your sampling?
20. How do you plan to determine the equivalence of groups?
21. If not all of those whom you solicit return questionnaires, what does this do to your sampling?

H. Studying Personality

1. Are you differentiating between manifest and covert personality trends?
2. To what extent is it possible to consider motives and unconscious purposes?
3. How do you propose to determine the values, attitudes, interests, motives, etc. that you plan to investigate?
4. Will interviews be biased if the interviewer has access to other information about an individual?
5. Can you depend on the subject's report to inform you about his motives, attitudes, problems, etc?
6. Are you sufficiently aware of the influence of dynamic mechanisms (repression, projection, denial, etc.) in the reports of your subjects?
7. How are you going to ensure that your subjects express their true feelings and attitudes?
8. If you plan to investigate underlying dynamics, are you prepared to establish close relationships with your subjects?
9. Is the questionnaire or testing approach going to permit you to get at the inner dynamics?
10. In a study using tests and other objective data, is it possible to question your subjects further to determine attitudes, beliefs, motives, etc?

11. Are you assuming that a verbal report of behavior is identical with the behavior?
12. Are you assuming that a verbal report of attitude is identical with the attitude?
13. Are you confounding a verbal report of memory with the actual facts?
14. Is the recall of childhood experiences a safe index of the actual experiences?
15. Which is preferable for your study—the questionnaire or interview approach?
16. In a study involving interviewing, have you considered the possibility of securing coöperation and frankness from your subjects?
17. To what extent is the interviewer influencing the answers to questions?
18. To what extent can you depend on self-testimony as compared with the testimony of others?
19. Are you assuming that inner expression of value, interest and feeling are the same as the judgment of a person made by some other person?
20. Is one justified in attributing dynamic significance to behavior for the purpose of research?
21. Do you prefer the free response type of inquiry (so-called open-ended questions) where the answers to questions must be categorized, or do you prefer recognition type questions where the answers can be directly tabulated?
22. In using observation material, how can you be sure that your observations are representative?
23. Have you made sure that the items which you have selected to represent a given trait are not biased in some other trait?
24. In thinking about personality characteristics, are these general or do they relate to specific situations?

I. Tests and Measures

1. Are you using the appropriate measures?
2. How is the material in your test to be selected?
3. Are your measures independent?
4. Have you taken into account the difficulty level of your test?
5. Do you wish to use a time-limit or work-limit?

6. Have you considered the relative merits of recall and recognition types of items?
7. Have you given consideration to the best type of score for your test?
8. How will you construct a scale to measure attitude?
9. Do you have a large enough reservoir of items to be able to make a selection of items in terms of difficulty, validity, etc?
10. How do you propose to make a composite of your measures? What would such a composite mean?
11. In using a questionnaire, do you propose to get a total score or merely to tabulate the answers to separate items?
12. To what extent are the responses of your subjects limited by the examiner or the directions?
13. Can you use a ready made measuring instrument or should you construct your own? If the latter, how will you determine its validity and reliability?
14. How valid are your measures?
15. Are there contaminating factors (age, sex, etc.) which might lessen the validity of the test you propose to use?
16. Do you propose testing the validity of your test on a new group?
17. Are your tests sufficiently reliable?
18. Do you propose to test the reliability of your measures?
19. Are you planning to use your time in scale making, testing validity and reliability when it could more profitably be directed elsewhere?
20. Are there norms with which to compare the findings in the group which you are proposing to study?
21. If you plan to use content material (like the T.A.T.) as a measuring device, have you specified the variables you propose to measure?
22. Have you properly disguised your intentions in the test you propose to use?
23. Should a pre-test be given as well as a test at the close of your experiment?
24. Is a pass-fail score on a test item identical with a qualitative score based on a content analysis?
25. Is a point score which is the sum of factors used in the solution to a problem adequate as a measure when there is a possibility that the factors may differ in importance?

26. If you plan to study the extremes on your scale, would it be well at the same time to pay attention to the intermediate points?

J. Use of Judgment

1. If you plan to use judgments, have you specified the basis on which your judgments would be made?
2. In securing ratings, whom will you select as your judges?
3. If you plan to use judgments, are you sure your judges have the necessary intelligence, information, background and other qualifications to permit them to make the judgments?
4. Can you depend on judgments to serve as a basis for selecting individuals as being mentally deficient, delinquent, schizophrenic, etc.
5. To what extent will subjective factors enter into judgments that you propose to make (or use) and how can these be avoided?
6. Which is better for your study—the overall judgment or a content analysis?

K. Content Analysis

1. How do you propose to determine your classification scheme?
2. Who is to make the classification scheme?
3. How reliable will the classification scheme be?
4. How do you propose to distribute material into the several classificatory categories?
5. Who is to distribute the items into categories?
6. Have you provided for determining the reliability of your classifications?
7. Have you sufficiently taken into account the subjective factor in analyzing content material?
8. Have you given attention to what sorts of behavior or responses will be included under such general categories as "resistance" or "rejection?"

L. Statistical Handling of Results

1. Do the conditions of your data warrant using the statistics which you propose to use?
2. Do your data satisfy the assumptions on which the statistical constants which you propose to use are based?

3. Have you enough cases to test fairly the significance of your statistical constants?

M. The Ed.D. Report

1. In preparing an Ed.D. project, have you considered the audience for which you will be writing?
2. In planning a Type B¹ project, are facts available concerning the situation in which the plan would operate?
3. In a Type B project, how will you determine the principles which will serve as the basis for your plan?
4. Have you given thought as to how you will defend your proposals in a Type B project?
5. In preparing a handbook, have you considered the manner of presentation?

¹ In Teachers College, Columbia University, a Type B project which may be submitted as partial fulfillment of the requirements for the Ed. D. degree is a plan for the improvement of education in some specific situation.

GENERAL RESPONSE PATTERN TO FIVE-CHOICE ITEMS

EDWARD L. CLARK

Northwestern University

Response tendencies of individuals to multiple choice test items have recently been the subject of several investigations. The tendency of some individuals to respond in a given direction, such as the selection of end or extreme responses in four or five choice items, has been studied by Berg and others (2, 3) to determine what association, if any, such tendencies may have with personality types or with clinical diagnoses. These studies have to do especially with the willingness of subjects to select end responses such as "dislike very much" or "like very much" and the evidence collected indicates that responses to such multiple choice items may be an individual expression of set.

Although tendencies to choose or avoid end responses have been observed in non-verbal tests of standard option form, similar studies in conventional multiple-choice tests of knowledge of subject matter or scholastic aptitude are rare. One hears informal opinions expressed occasionally to the effect that students neglect the third position or the fourth position of five-choice tests, and more rarely such opinions get into print. Atwell and Wells (1) state that "where the answers are pure guesses the position of misleads is a distinct factor in the ones which are selected. On such a selection a definite pattern is found, positions one, five, three, four, two, being chosen with decreasing frequency in that order." No supporting data accompany this assertion. McNamara and Weitzman (6) find that items having right answers in the fourth position are the most difficult, those with right answers in the second and third positions are the easiest, and the first and fifth positions are of equal intermediate difficulty. Their results for four-choice items show that the third position is the most difficult, and this finding is interpreted as agreeing with the results for five-choice items in that the next-to-last position is the most difficult. Otherwise, there seems to be little agreement between findings for positions for five-choice and four-choice items. To exclude the possibility that item writers, responding inadvertently to some common belief, had placed the

more difficult right answers in the next-to-last position, a comparison was made between such items in one form of tests with the same items in other forms where the right answer was not in the next-to-last position. This comparison did, indeed, show that items were slightly easier when the answers occurred in other than the next-to-last position. These investigators did not make this definitive test for any of the other positions. Cronbach (5) expresses the opinion that the five-choice type of item is, as compared to the two-choice item, quite free from the influence of position of the right answer.

The purpose of this study is to investigate more fully than has been done in previous studies a possible tendency among college students to prefer certain response positions to the neglect of others in five-choice tests. The preliminary findings were collected during the eight-year period when the Clark Analogies Test was being constructed (4). Because both the right answers and the wrong answers were placed by casting dice, it was hoped that no repeating pattern of right answers nor special placement of attractive distractors characterized the early editions of this test. However, the relatively small number of items used—about five hundred during the eight years—did not preclude the possibility that the rare distractor of great attractiveness should happen to occur more often in one position than another. Consequently it was deemed necessary to collect distributions of wrong answers from three additional tests: N. U. Form A, a scholastic aptitude test made of two hundred difficult word definitions; three forms of the Ohio State University Psychological Test; and three subtests of The Guilford-Zimmerman Aptitude Survey.

The data in Table I are based, with the exception of Experimental Analogies I and II, on the wrong answers made by applicants for admission to Northwestern University. The entries of the table give the ratio of observed to expected numbers of wrong answers for each of the five positions of the several tests. To determine the expected number for each position, one-fourth was taken of the wrong answers to all items having the right answers in the other four positions. For example, the expected number of wrong answers in the first position was one-fourth of all wrong answers given to items whose right answers were in positions two, three, four, or five. Because the total number of wrong answers was not the same for each set of items or the expected number of wrongs in each position was not always just the same, the un-

TABLE I—DATA ON THE DISTRIBUTION OF WRONG ANSWERS IN THE FIVE POSITIONS OF MULTIPLE CHOICE TEST ITEMS

Name of Tests	No. of Items	Ratios of Observed to Expected Numbers of Wrong Answers for the Five Positions					No. of Wrong Answers	Approx. No. of Examinees
		1.	2.	3.	4.	5.		
Clark Analogies	500	1.1010	1.0290	1.0136	0.9895	0.8692	92,822	2000
N.U. Form A	200	1.0485	1.1681	1.1070	0.8044	0.8560	41,531	400
Ohio Psychological	450	1.0866	1.0532	1.0080	0.9496	0.9028	44,776	640
Guilford-Zimmerman	146	1.0406	1.0504	1.0940	0.9923	0.8233	42,577	1200
Exp. Analogies I	(30)	0.9816	0.9913	1.0009	1.0224	1.0099	5,224	628
Exp. Analogies II	(55)	1.0974	1.0174	1.0001	0.9112	0.9691	11,423	400
C. Anal. Forms A & B:								
Given first	(300)	0.9964	0.9996	0.9835	1.0429	0.9814	15,026	272
Given second	(300)	1.0736	0.9973	0.9543	1.0134	0.9600	15,021	(272)
Totals	1,296	1.0703	1.0538	1.0335	0.9568	0.8837	268,400	5,540

weighted average of the five ratios of observed to expected wrong answers for each test will not necessarily be 1.0000.

RESULTS

The results shown on the first line of Table I are based on the numerous early editions of the analogies test in which approximately five hundred different items were tried out in tests no longer than one hundred and fifty items each. The ratio, 1.1010, of observed to expected number of wrong answers for position one was based on the twenty thousand, five hundred and seventy-one wrong answers made in this position to the eighteen thousand, six hundred and eighty-four wrong answers expected: the expected eighteen thousand six hundred and eighty-four is one-fourth of all wrong answers given to all items having right answers in position two, three, four, and five. The remaining ratios of observed to expected number of wrong answers shows a declining sequence. In the five positions, ninety-two thousand, eight hundred and twenty-two wrong answers were tabulated from the papers of approximately two thousand examinees. A chi-square based on the five pairs of observed and expected wrong answers was found to be five hundred and thirty-nine or about forty times the value needed to be significant at the one per cent level of confidence. In the second line, the

ratios depart from 1.0000 in a very significant manner but they do not exactly parallel those of the first line.

Attention is called to the fact that N. U. Form A was a very hard test of two hundred items. The four hundred examinees all took the entire test and on the average each made over one hundred wrong answers. The results on the third line are based on the three editions of the Ohio State University Psychological Test of one hundred and fifty items each and the ratios obtained depart from 1.000 in a manner similar to that found for the analogies tests. The data of the fourth line, derived from the three short subtests of the Guilford-Zimmerman battery, again show ratios, departing considerably from 1.000, which are statistically stable.

Although the first four groups of tests did agree in showing relatively more wrong answers in the first positions than in the last—a tendency of great statistical stability—each test differed significantly from each other one. To understand this lack of agreement in results among the tests, it is necessary to examine the possible effects of the placement of unusually attractive misleads. For most of the tests used in this study the average item was answered correctly about sixty per cent of the time. Thus, the four wrong options each got an average of about ten per cent of all the responses. There were, however, a few items which had very attractive misleads which alone drew responses from fifty or even sixty per cent of the examinees. Even when the position of all options is determined by dice, there is little reason to expect that the relatively small number of items with very attractive misleads will be evenly distributed in the five positions of any one test. Getting tens of thousands of wrong responses from examinees only serves to establish with great stability the less than symmetrical distribution of attractive misleads. The results obtained for the N. U. Form A test were especially subject to the chance distribution of attractive misleads. Since popular misconceptions of the meanings of words were frequently used as wrong options in this test, every eighth or tenth item had a mislead which was used more often than the right answer. Much of the lack of agreement among the first four groups of tests as to the relative attractiveness of the five positions must be due to differences in the chance distribution of the very attractive wrong answers.

In order to control rigidly the chance placement of very attractive wrong options, Experimental Analogies I was designed. Thirty

items were used, six having right answers in each of the five positions, and four forms of the test were made. The right answer of each item always occurred in a given position but in the four forms the items had slightly different order on the page and the position of the wrong options were systematically varied according to a Latin square: each wrong option occurred once in each position and always followed a different wrong option.

The results of Experimental Analogies I when given to six hundred and twenty-eight students of elementary psychology are shown in line five of the table. The differences between observed and expected numbers of wrong answers for this one-page test were amazingly slight: a chi-square test of the data indicated that a random drawing of five thousand, two hundred and twenty-four cases of five equally probable events would in over ninety-five cases out of one hundred show greater departures than our obtained ratios. To assume any tendency of these six hundred and twenty-eight students of elementary psychology to give preference to any of the five positions of this test would be entirely unwarranted.

In order to explore further this remarkably even distribution of wrong answers, the following year a longer experimental analogies test having fifty-five items was constructed in four forms using varied order of items and, again, a Latin square design for the placement of the wrong options was used. The results from this fifty-five-item test administered to four hundred elementary psychology students gave the ratios shown in the sixth line. Although the Latin square design was used just as it had been used the year before, the departures from chance of the ratios expected were highly significant (chi-square was three times the value required for significance at the one per cent level) and resembled roughly those departures from chance which had been found earlier for the four standardized tests. The neglect of position four to be observed in Experimental Analogies II was found, upon examination of the data, to be due entirely to what happened for nine of the eleven items which had their right answers in position three. That is to say, the eleven items having right answers in position one, the eleven having right answers in position two, and the eleven with right answers in position five did not show a neglect of wrong answers placed in position four.

The results from the two experimental analogies tests are not

in agreement with each other although both used the Latin square design for the position of wrong options. All conditions for the collecting of data for the two tests were alike, apparently, except for the number of items in the tests. Nothing was said at either testing about a need for haste in finishing the test, but the fact that the first test was on one page and the second test was on two pages must have affected the attitude of the subjects towards it. Although the class period provided ample time to do all fifty-five items, an occasional student must have felt hurried and must not have read beyond the option which he considered the right one.

Since the results from Experimental Analogies II suggested that a feeling of haste tended to make examinees neglect positions four and five, it was reasoned that subjects might also neglect the last positions because of fatigue or boredom. To test this hypothesis, data were used which had been gathered from two hundred and seventy-two applicants who had taken Form A and Form B at one testing session. One-half had taken Form A first and one-half Form B first; all took both tests with a ten-minute rest interval between the two ninety-minute periods of work. Since both Form A and Form B occurred an equal number of times as the first test (and as the second test), all the attractive wrong answers could have occurred in one position, e.g., in position four, without invalidating the comparison of the results from first and second testings. The chance placement of attractive wrong answers was exactly the same in the two periods of testing and, thus, ruled out of this comparison. The third line from the bottom of the table of data shows the use of the five positions for wrong answers when either Form A or Form B was the first test taken, and the next line below shows what happens when either of these two tests was the second taken. It will be noted that the total number of wrong responses was almost exactly the same on second taking as it was on the first: any practice effects must have been hidden by the results of fatigue or boredom. (The r between the two forms was 0.92) The difference in the use of the five positions for wrong answers on taking the second form of the test, as contrasted with the first form, were as follows: the first position was used two hundred and forty-two more times and all the remaining positions were used fewer times—eighty, ninety, eighty-three, and sixty-six, respectively. Thus, the first position was used much more often on the second taking of a form of the

test and the last three positions were used less often. A chi-square test for these differences indicated significance at the two per cent level of confidence.

In summary, the great mass of data of the study which is based on the four standardized tests (lines one to four in the table) indicates that there was some neglect of the last positions in five-option multiple choice items. The lack of complete agreement between tests as to positional preferences of students is probably due to the chance placement of the relatively few very attractive misleads or wrong options. Data from the relatively small groups of subjects (lines five to eight of the table) suggest that the pressure of time limitations and fatigue or boredom are effective in making a declining frequency of the use of the five positions.

The general results of this study, as summarized by the last line of data, for all tests, show a tendency of a few subjects to neglect the last positions—a failure to read all the way through an item before making a response. To assume that once in one hundred items students read only the first two options, once in a dozen items they read but three options, and once in twenty or twenty-five they read but four options would give results much like those in the last summary line of the table. The data suggest that under usual testing conditions with a time pressure and, in long tests, a certain amount of boredom, students will in less than ten per cent of the items read the first three options only, in very rare cases they read only two options and again rarely, they fail to read the last option.

CONCLUSIONS

- (1) General positional preferences as they exist with five-choice items are weak.
- (2) When the relative attractiveness of misleads is experimentally controlled and when examinees are under no time pressure to finish, positional preferences are, indeed, infinitesimal.
- (3) Working under the usual pressure of time, subjects use the five positions in a slightly declining frequency.

REFERENCES

- 1) C. R. Atwell and F. L. Wells, "Wide Range Multiple Choice Vocabulary Tests," *Journal of Applied Psychology*, 21: 550-555, 1937.
- 2) Irwin A. Berg, "The Reliability of Extreme Position Response Sets in Two Tests," *Journal of Psychology*, 36: 3-9, 1953.

- 3) Irwin A. Berg and J. S. Collier, "Personality and Group Differences in Extreme Response Sets," *Educational and Psychological Measurement*, 13: 164-169, 1953.
- 4) Edward L. Clark, *Clark Analogies Test, Forms A & B*. Evanston, Illinois, 1953.
- 5) L. J. Cronbach, "Response Sets and Test Validity, *Educational and Psychological Measurement*, 6: 475-494, 1946.
- 6) W. J. McNamara and E. Weitzman, "The Effect of Choice Placement on the Difficulty of Multiple-Choice Questions," *Journal of Educational Psychology*, 36: 103-113, 1945.

STIMULUS FAMILIARIZATION AS A FACTOR IN IDEATIONAL LEARNING¹

WALTER WEISS and BERNARD J. FINE

Boston University

A variety of research has indicated that the association of responses with stimuli is facilitated by prior familiarization with the stimuli. For example, Lawrence found that the learning of a new instrumental response is faster with a familiar cue than with one to which the animal had not learned to attend (3, 4). Other investigators have demonstrated that the prior association of irrelevant responses with critical stimuli facilitated the later learning of the desired responses with those stimuli (1, 5, 6). Stimulus familiarity has also been shown to facilitate serial-order learning (2) and tachistoscopic recognition (7). While humans have been used as subjects in most of the studies, the problems presented were not complex ones and did not require much use of the implicit symbolic mechanisms called for in learning ideational material from mass media.

The main objective of the present research was to test the prediction that prior familiarization with key terms will facilitate learning from an informational communication. A second purpose is related to a pragmatic consideration. Since the familiarization procedure takes time, is it more advantageous to expose the learners twice to the communication without prior familiarization than once preceded by familiarization training?

METHOD

Experimental Conditions and Materials. Three groups were formed comprising a familiarization-training group, a one-showing control group, and a two-showing control group. Hereafter, they will be designated respectively as *E*, *C1*, and *C2*. All three were exposed to an edited version of a film-strip on the organization of the United Nations, called "Structure for Peace." The structures and functions of the six main subdivisions of the UN were de-

¹ This research was supported by the United States Air Force under Contract No. AF 18(600)-1210, monitored by the Training Aids Research Laboratories, Chanute Air Force Base.

scribed in some detail in a recorded script which was synchronized with the film-strip. For *C1* this completed the learning aspect of the session. *C2* was immediately re-exposed to the communication. The *E* group saw the film-strip only once but immediately prior to that underwent a familiarization-training procedure. It was only *after* completing this procedure that the *E* classes were informed that they were to view a film-strip on the UN.

The visual material used in the familiarization procedure consisted of six of the twenty-four content-related frames of the prepared film-strip. Each of the frames was a drawn representation of a major subdivision of the UN but was bare of any identifying material.² A paired-associates procedure modified for group presentation comprised the familiarization technique. As each frame was shown, a recorded voice indicated the correct identifying name for that subdivision. The *Ss* simply listened and learned during this sequence. Then, the frames were presented again, but without the recorded identification of the names. This time the *Ss* had eight seconds in which to select the name of the appropriate organ from among the six printed on a prepared sheet. They did this by placing the number 1 next to the judged name for the first frame shown, etc. The recorded voice called out the number of each frame as it was shown. This arrangement of a listening trial and then an answering trial was repeated six times.³ The order of the frames was not changed from a listening to an answering trial; but from one listening-answering sequence to the next, the order was changed in accordance with a predetermined six by six Latin square arrangement. The familiarization training took eleven minutes, which was equal in time to that needed for one showing of the informational film-strip. Thus, both the *E* and *C2* groups spent the same amount of time in the learning aspect of the session.

Before the film-strip was presented, the *Ss* were told that they would be tested immediately afterwards on its content. The after-questionnaire contained twenty multiple-choice items related to the structures and functions of the six main parts of the UN and

² When a subdivision of the UN was described in the film-strip, the relevant frame from among the six used in the familiarization procedure was always the first one shown.

³ Approximately eighty per cent of the *Ss* in the *E* classes responded without error on the last *two* answering trials and approximately ninety per cent were correct on the last trial.

five additional questions on the UN covering general knowledge not presented in the film-strip. These five were designed to provide a rough estimate of the Ss' pre-existing levels of information about the UN. A few opinion items on reactions to the film-strip were also included.

Subjects. Approximately seven hundred upperclass students from three high schools in the Boston area were tested. Since intact classes were used, the N for the statistical analyses is based on classes as the sampling units. Eleven were assigned to the E group, ten to the $C1$ group, and seven to the $C2$ group. Two or three classes were examined during each school hour. When two classes were tested, they were assigned randomly to the E and $C1$ groups; otherwise they were assigned at random to the three treatments. One $C1$ class was eliminated as a disciplinary problem. The number of Ss in each class varied from fourteen to thirty-four.

RESULTS

Before examining the effect of the treatments on learning from the film-strip, it is important to determine if differences existed between the groups in their prior fund of knowledge about the UN. A rough index of knowledgeability is given by the mean scores on the general information questions in Table I. None of the differences between the groups is significant.

Table I also contains the means for each group on the twenty fact-quiz items related to the content of the film-strip. Significant differences are obtained between the E and $C1$ groups ($p < 0.05$, one tail) and between the $C2$ and $C1$ groups ($p < 0.05$, two tails). Furthermore, these differences hold for the "easy" questions as well as the "difficult" ones. The means of the E and $C2$ groups do not differ reliably from each other ($p = 0.50$, two tails).

TABLE I—MEAN NUMBER OF FACT-QUIZ ITEMS CORRECT

Group	N	General Information about UN		Content of Film-Strip	
		Mean*	SD	Mean†	SD
E	11	2.43	0.77	11.75	2.09
$C1$	10	2.15	0.60	10.38	1.28
$C2$	7	2.31	0.92	12.41	0.92

* Based on five questions.

† Based on twenty questions.

TABLE II—JUDGMENTS OF INTEREST VALUE OF FILM-STRIP AND OWN MOTIVATION TO LEARN

A. Mean per cent per group judging the film-strip "very interesting" or "somewhat interesting."

Group	N	Per cent	SD*
<i>E</i>	11	50.4	9.04
<i>C1</i>	10	62.6	8.95
<i>C2</i>	7	57.8	9.03

B. Mean per cent per group who "tried very hard to learn" or "tried somewhat hard to learn" while attending to the film-strip

Group	N	Per cent	SD*
<i>E</i>	11	50.8	9.68
<i>C1</i>	10	62.0	13.60
<i>C2</i>	7	62.5	14.23

* Based on arc sine transformation.

Since different fact-quiz items tested knowledge of the structures and functions of the different parts of the UN, the mean of each group of questions related to a particular subdivision was computed. For each subset of questions, the mean for the *E* group is higher than the corresponding one for the *C1* group. The differences are reliable for three of the six means. The same results are obtained for the comparisons between the means for the *C2* and *C1* groups. While the *C2* group is slightly superior to the *E* group on all of the subsets of questions, the differences between any pair of means do not approach significance.

The after-questionnaire contained two opinion items on the *Ss'* reactions to the film-strip. One requested them to indicate how interesting they thought the film-strip was; and the other asked them to indicate their "felt" motivation-to-learn while attending to the film-strip. The data in Table II reveal that the percentages of *Ss* in the familiarization group who considered the film-strip interesting or who stated that they tried "hard" to learn were smaller than the corresponding ones for the *Ss* in the *C1* and *C2* groups. However, none of these differences reach the 0.05 level, two tails.

DISCUSSION

The results of the research confirm the prediction that prior familiarization with key terms will facilitate learning from an informational communication. Furthermore, the time devoted to the

Pre-Training on Subsequent Motor Performance," *Journal of Experimental Psychology*, 46: 329-336, 1953.

6) I. L. Rossman and A. E. Goss, "The Acquired Distinctiveness of Cues: the Rôle of Discriminative Verbal Responses in Facilitating the Acquisition of Discriminative Motor Responses," *Journal of Experimental Psychology*, 42: 173-182, 1951.

7) R. L. Solomon and D. H. Howes, "Word Frequency, Personal Values, and Visual Duration Thresholds," *Psychological Review*, 58: 256-270, 1951.

BOOK REVIEWS

HERBERT A. THELEN. *Dynamics of Groups at Work*. Chicago: University of Chicago Press, 1954, pp. 366. \$6.00.

Since the death of Lewin in 1947, most investigation of group processes has consisted of differentiation and exploration of cells within larger frameworks, and perhaps rightly so. The extravagant promise of much of the early exploration has given way to sober work on focused problems. However, this has meant that relatively less attention has been given to comprehensive theory-building.

The author of this book has characteristically been interested in integrative concepts, and has made periodic appearances in the literature with pieces of what promised to be a useful set of notions about group life. This book presents such a framework, drawn from eight years' work at the Human Dynamics Laboratory at the University of Chicago.

The approach is essentially molar, drawing systematic empirical conclusions from experience in specified settings (e.g., classroom teaching, training, community action), and developing explanatory notions to unify these. The book can be said to be non-Lewinian—in the topological sense—although the author's debts to Lewin are clear. It is probably one of the best over-all treatments of group life now available. Thelen has drawn most directly on Deweyan problem-solving and general psychiatric theory, notably that of W. R. Bion, to produce an approach highly integrative of the "thought-feeling-action" trichotomy. A weakness of Dewey's was his failure to understand the psychiatric dimensions of interpersonal transactions; the weakness of his followers has been a tendency to ideological preachment. This book avoids both these traps, most clearly the former.

The first half of the book deals with "technologies"—sets of "principles useful to bring about change toward desired ends." Successively, community action groups, classroom teaching, faculty in-service training, administrative decision-making, group development training, and large public meetings are examined, and proposed models are presented for each of these, often with considerable freshness and ingenuity.

The second section presents conceptual material which serves to explain and amplify the material in the first section, and relate

it to a larger framework. The chapters deal with the individual as he functions in groups ("membership," and the thought-feeling-action relationship); the reality factors within and surrounding the group; the development of group control systems; leadership; and (a severely neglected area in this field) the community as the "context of group operation."

A number of recurrent themes appear: that experimental (feedback-using) method is most desirable; that group behavior represents a mixture of work and emotionality—the result of member attempts to work on private problems, build and maintain a group, and meet the demands of an objective task; that the notion of "membership" is crucial for understanding anyone's behavior; and, finally, that we must think of the "reality-centered" group, working autonomously to meet intra-member, intra-group, and inter-group demands.

Many useful and relatively novel ideas are presented. The notion of the "bridging" group as a means of decreasing inter-group conflict is one, as is the ingenious model of organizational decision-making which successfully separates line and staff ("pressure" and "coöperation") functions, while integrating task effort. The "principles underlying the control of the group," in Chapter 10, comprise an image of group operation that has not had wide currency; the case study of the trainer rôle is equally provocative. The community-centered model of faculty in-service training, the description of community-level processes of communication and communion, and the discussion of pressure-oriented and coöperation-oriented groups in the community promise to be useful in correcting the current tendency to focus only on intra-group processes. Finally, the book reminds us constantly, between the lines, that all groups have a training dimension, and that improvement of group operation may consist largely of enlarging and legitimizing this dimension.

The book (as the author wryly remarks at the end) is honey-combed with lists of generalizations and principles. Some of these, such as the suggestions for composing subgroups, the ideas on transfer of training, and the discussion of inter-organizational conflict, are now in directly testable form. Others are parts of systematic models, and would require considerable translation before assuming the form of hypotheses.

The style is compact, frequently abstract, rarely unnecessarily

complex (as in some earlier publications by the author), often epigrammatic, and sometimes funny. The cross-referencing is extremely thorough, and there is a good index.

This is a needed book. One might wish for more references to the molecular-level work done at the Laboratory; and readers who have not seen the Human Dynamics Laboratory monograph on *Methods for Studying Work and Emotionality in Group Operation* (1954) may complain of the imprecision of the book. It is certainly true that the parts of the general theory presented, while internally well related, are by no means parsimonious. In a real sense, this book approaches the limits of what can be done, theorywise, with ordinary language, and it is possible that the direction from here lies in increasingly abstract and mathematical formulations. Finally, careful testing and reconstruction of the theory presented is as much needed now as was integrative, molar work prior to the publication of this book.

MATTHEW B. MILES

Teachers College, Columbia University.

PAUL WITTY. *How to Become a Better Reader*. Chicago: Science Research Associates, Inc., 1953, pp. 304.

This text is designed to help people, in school or in work situations, to improve rate of reading, comprehension, and joy of reading. It is organized as a self-teaching book. The first part consists of twenty lessons. This is followed by a classified bibliography of books the person may want to read. The last part of the book is made up of twenty exercises with tests. One of these is to follow each lesson in the first part of the book.

The early lessons should provide motivation to improve reading. They are concerned with benefits of reading, possibility of improvement, needs, nature of eye-movements, and reading for a purpose. The subsequent lessons deal with improving various reading skills such as rate, skimming, finding the main idea, details, evaluation, improving vocabulary, study-type reading, and how to keep on reading better.

Each lesson and accompanying exercise is carefully and effectively organized so that the reader knows just how to proceed and how to measure progress in speed, comprehension and vocabulary knowledge. The exposition is clear, and the reading selections are well-chosen.

The book provides a means for the person who is strongly motivated to help himself to better reading. Although a self-help text, it is probable that best results will be obtained when it is used with supervision of a reading specialist.

MILES A. TINKER

University of Minnesota

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

March, 1956

Number 3

FACTOR MODELS

A. E. MAXWELL

Department of Psychology, Maudsley Hospital, London.

Psychologists in general and factor analysts in particular are glad to learn that the advent of electronic computers soon may lighten the burden of their calculations. Consequently the recent article in this journal by Wrigley and Neuhaus (9) on the use of an electronic computer in principal axes factor analysis (3) is of special interest, and is likely to be read widely. For this reason some remarks they make in their opening paragraphs—presumably by way of vindicating the principal axes method in the eyes of psychologists—should hardly be allowed to pass without comment.

TWO MODELS

The writers appear to rank the principal axes method and the centroid method of analysis together simply as different methods within a single model. "The advantage of the centroid method," they say, "has been that it has given a reasonably good approximation to a principal axes factor analysis . . ." Now it would be fatuous at this late date to re-open the old controversy between the advocates of these two methods, but it would be careless to gloss over points of fact concerning them. Two different models are involved, the principal axes model of Pearson and Hotelling (3) on the one hand and the centroid method of Thurstone (8) on the other. Letting p stand for the number of tests and n for the size of the sample then as Kendall (4) puts it, the principal axes model considers n points in p dimensions, while the centroid model considers p points in n dimensions. The "loss of orthogonality" between the centroid axes to which Wrigley and Neuhaus refer is not an indication of any inferiority inherent in the centroid model, but rather a demonstration that the arithmetic techniques so far available for use with this model are mathematically inferior to those available for use with the principal axes model.

Another difference, emphasised by Bartlett (1), is that whereas the principal axes method is an empirical one and no psychological hypotheses are involved, the centroid method is generally employed with "communalities" in the diagonal cells of the matrix to meet the psychologist's very reasonable claim that the correlations (or covariances) between a set of tests can be described most meaningfully by a few "common" factors, together with a factor specific to each of the several tests concerned. A useful discussion of the point can be found in an article by Burt (2) in which alternative methods of factor analysis are considered.

Now, although the psychologist need not worry too much about the mathematical details to which reference has just been made, he should be alert lest a mathematical model incompatible with his psychological hypothesis gets sold to him on its arithmetic merits alone. He should consider whether he is justified in renouncing his birthright as a *factor* analyst at this or any price. While the writer has much appreciation of the principal axes method of "breaking down" a correlation or covariance matrix, he is not convinced that it is in general suited to the psychologist's needs.

LIMITATIONS OF THE PRINCIPAL AXES METHOD

The principal axes method too—for whatever purpose it is used—has its limitations. These have not been stated by Wrigley and Neuhaus although they are at pains to list the advantages of the method. First, the loadings obtained are not invariant under changes of scale in the variates (4, pp. 65–66); values got if the covariance matrix instead of the correlation matrix is employed are not proportional. Thomson (7, p. 335) points out that such invariance is a necessary requirement of any satisfactory factor method. Second, apart from tests for the significance of successive latent roots by Kelley, and by Bartlett (to whom they refer), the authors omit to point out that we still have insufficient knowledge about the sampling issues involved in the principal axes procedure to enable us to test the significance of loadings obtained by its use, or to test the difference between loadings obtained on different occasions. Here one must add that the latter limitation applies also to the centroid method; nor have we a valid test of the significance of centroid factors, but "simple structure," when present, is invariant under changes of scale in the variates.

ANOTHER APPROACH

In view of what has been said it would seem desirable to seek another approach which would satisfy the psychologist's needs and at the same time reasonably fulfill optimal statistical requirements. With the aid of an electronic computer the "maximum likelihood method" of Lawley (5; 7) is now a possible choice. Earlier doubts about its mathematical and statistical validity, except when the number of tests is very small, have not been confirmed and this method, Bartlett tells us (1, pp. 27-28), should meet our requirements. There may be good reasons on the programming side why it has not been chosen by Wrigley and Neuhaus, but if there are not it offers many attractions. It gives results which are invariant under changes of scale in the variates. This is indeed an advantage, for loss of information which otherwise would result from any use of standardized scores can be obviated. It provides too a valid statistical test of the number of significant factors, and last, since the sampling theory of the method is now well under way (6), reliable comparisons between results obtained from different studies will soon be possible. Admittedly the method, by making allowance for "specific" factors, shares the limitation of the centroid method (with communalities) that factor scores can be "estimated" only. But factor analysts seldom employ factor scores, and the need to do so will be even less when loadings can be compared by valid statistical methods and differences tested for significance.

SUMMARY

The geometrical distinction between the principal axes model and the centroid model is pointed out, and attention is drawn to the special way in which the factor analyst is wont to use the latter. Some statistical limitations are then mentioned and an alternative approach suggested.

REFERENCES

- 1) M. S. Bartlett, "Factor Analysis in Psychology as a Statistician Sees It," *Nordisk Psykologi's Monograph Series* No. 3: 23-34, 1953.
- 2) C. Burt, "Alternative Methods of Factor Analysis," *British Journal of Psychology, Statistics Section* 2: 98-121, 1949.
- 3) H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, 24: 417-441, 498-520, 1933.

4) M. G. Kendall, "Factor Analysis," *Journal of the Royal Statistical Society*, (series B), 12: 60-73, 1950.

5) D. N. Lawley, "The Estimation of Factor Loadings by the Method of Maximum Likelihood," *Proceedings of the Royal Society, Edinburgh*, 60: 64-82, 1940.

6) D. N. Lawley, "A Modified Method of Estimation in Factor Analysis and Some Large Sample Results," *Nordisk Psykologi's Monograph Series* No. 3: 35-42, 1953.

7) G. H. Thomson, *Factorial Analysis of Human Ability*, fifth edition. London, University of London Press, 1951.

8) L. L. Thurstone, *Multiple Factor Analysis*. Chicago, University of Chicago Press, 1947.

9) C. Wrigley and J. O. Neuhaus, "The Use of an Electronic Computer in Principal Axes Factor Analysis," *Journal of Educational Psychology*, 46: 31-41, 1955.

AMERICAN DOCTORAL DISSERTATIONS ON EDUCATIONAL PSYCHOLOGY IN FOREIGN COUNTRIES

WALTER CROSBY EELLS

Washington, D. C.

The writer has recently completed a study of some fifteen thousand doctoral dissertations in the field of education which have been accepted by American institutions of higher education. More than one thousand of these have dealt with various phases of education in foreign countries.¹

Of the dissertations concerned with foreign education, thirty-two deal specifically with some aspect of educational psychology in various countries. These are spread over a period of seventy years, the earliest having been written at Johns Hopkins University in 1884. Interestingly enough it was written by John Dewey on the "Psychology of Kant." Almost half of them, however, have been written during the past decade.

These dissertations have been accepted at twelve American universities; Columbia leads with nine, followed by New York University with seven. Chicago is credited with four, and Johns Hopkins, Ohio State, and Southern California have two each. A half dozen other institutions have one each.

Following is a list of the dissertations on educational psychology in foreign countries with name of author, title of dissertation, institution at which it was accepted, and date of acceptance.

¹ Titles of all dissertations in education and related fields have been examined in *Doctoral Dissertations Accepted by American Universities* (New York: H. W. Wilson Co., 1934-54, 21 vols.); *Bibliography of Research Studies in Education* (Washington: United States Office of Education, 1929-41, 13 vols., and continued in manuscript to 1952); W. S. Monroe's *Theses in Education Accepted by American Colleges and Universities* (Urbana: University of Illinois, 1920-28, 6 vols.); *American Doctoral Dissertations Printed* (Washington: Library of Congress, 1912-38, 27 vols.); and many other specialized bibliographies, catalogs, and abstracts published by individual institutions and organizations. For a brief report on the study see W. C. Eells, "American Doctoral Dissertations on Foreign Education," *Higher Education*, 12: 19-22, October 1955.

GENERAL

Casanova, Teobaldo. "Educational Psychology and Some Aspects of Education in Latin America." New York University, 1933.

Spaulding, Seth J. "An Investigation of Factors which Influence the Effectiveness of Fundamental Education Reading Materials for Latin-American Adults." Ohio State University, 1954.

Terman, Earle Luther. "The Development and Application of National Educational Survey Techniques with Special Emphasis upon Criteria for Measuring Intelligence Internationally." New York University, 1930.

EUROPE

General

Klineberg, Otto. "A Study of Psychological Differences Between Racial and Nationality Groups in Europe." Columbia University, 1931. (Includes groups from France, Germany, and Italy.)

Spencer, Margaret. "Traits of Notable Educators." New York University, 1937. (Includes educators from Czechoslovakia, France, Germany, Great Britain, Greece, Italy, and Switzerland.)

France

Lerner, Ruth Spero. "Dr. Alfred Binet's Contributions to Experimental Education." New York University, 1933.

Germany

Cole, Percival Pichard. "Herbart and Froebel: An Attempt at Synthesis." Columbia University, 1907.

Dewey, John. "Psychology of Kant." Johns Hopkins University, 1884.

Holaday, Beverly Eli. "Educational Psychology in the Third Reich." Ohio State University, 1937.

Liebman, Rebekah R. "Herbartian Methodology in American Education." Johns Hopkins University, 1949.

Meyer, Adolph Erich. "William Stern: His Contribution to Experimental Education." New York University, 1926.

Siegmeister, Walter. "Theory and Practical Significance of the Experimental Pedagogy of Dr. Rudolf Steiner." New York University, 1932.

Great Britain

Aleck, Adolph William. "The Essentials of Mental Hygiene in the History of Education to Herbert Spencer." New York University, 1931.

Italy

La Malfa, Joachim J. "Educational Psychology in Italy." University of Michigan, 1949.

O'Reilly, Charles T. "Race Prejudice Among Catholic College Students in the United States and in Italy: A Comparative Study of the Role of Religion and Personality in Inter-Group Relations." University of Notre Dame, 1954.

Spain

Daly, Walter Aloysius. "The Educational Psychology of Juan Luis Vives." Catholic University of America, 1924.

ASIA

General

Leiter, Russell G. "A Comparative Study of the General Intelligence of Caucasian, Chinese, and Japanese Children, as Measured by the Leiter International Performance Scale." University of Southern California, 1939.

China

Chan, Ying. "The Development of Parallel Reading Comprehension Examinations in English and Chinese at the Graduate Level." Columbia University, 1954.

Chang, Chung Yuan. "A Study of the Relative Merits of the Vertical and Horizontal Lines in Reading Chinese Print." Columbia University, 1943.

Ho, Perry Yewton. "A Comparative Study of the Ingenuity of American and Chinese Students." University of Southern California, 1926.

Hu, I. "An Experimental Study of the Reading Habits of Adult Chinese." University of Chicago, 1928.

Lew, Timothy Ting Fang. "The Psychology of Learning Chinese: A Preliminary Analysis by Means of Experimental Psychology of Some of the Factors Involved in the Process of Learning Chinese Characters." Columbia University, 1920.

Liu, Herman Chan En. "Non-Verbal Intelligence Tests for Use in China." Columbia University, 1922.

Shen, Eugene. "The Reading of Chinese in Vertical and Horizontal Axes: An Experimental Study by Means of Photographic Recording of Eye Movements." Stanford University, 1926.

Shen, Ye. "A Proposed Program for a Chinese Junior High School in Respect to Gifted Children as Compared with Other Children." Columbia University, 1936.

Tu, Horace Tsou Chow. "The Effects of Different Arrangements of the Chinese Language upon Speed and Comprehension of Silent Reading." State University of Iowa, 1924.

Wang, Fung Chiai. "An Experimental Study of Eye-Movements in the Silent Reading of Chinese." University of Chicago, 1933.

India

McGavran, Donald Anderson. "Education and Beliefs of Popular Hinduism: A Study of the Beliefs of Secondary School Boys in Central Provinces, India, in Regard to Nineteen Major Beliefs of Popular Hinduism." Columbia University, 1936.

Iran

Boyce, Arthur Clifton. "The Comparison of Two Forms of Persian Writing by Means of Scales Constructed and Standardized for the Purpose." University of Chicago, 1933.

Japan

Otomo, Shigeru. "An Experimental Study of the Eye Movements Made by Various Persons in Reading Japanese Texts of Different Forms." University of Chicago, 1924.

Philippines

Carreon, Manuel Lingad. "Applicability of Standard Tests to the Filipino." University of Minnesota, 1923.

Mendez, Paz P. "Family Relations: A Test for Filipino High School and College Youth." Columbia University, 1952.

DIFFERENCES BETWEEN GIRL AND BOY "REPEATERS"

WILLIAM W. WATTENBERG

Wayne University

It has been generally assumed that the forces responsible for delinquency operate in much the same fashion for boys and for girls. True, the offenses which are entered upon the police and court records are different; girls in higher proportion are charged with incorrigibility, sex offenses, and truancy from home. This fact can easily be attributed to cultural determinants, which influence the sensitivity of the community to these types of misconduct in girls and which impel girls towards these forms of antisocial conduct in contrast to boys whose delinquencies run more to thefts and assaults.

A study, just completed, by Wattenberg and Saunders (1), upon factors linked to repeating among girls indicates that in some respects girls are led to repeated offenses by a somewhat different combination of forces than holds true for boys. In that study, the records of all girls contacted on charges of misconduct by the Youth Bureau of the Detroit Police Department in 1952 were compared in accordance with whether or not the girl had a previous record. All items were tabulated and the differences tested by the chi-square technique. The pattern of items where the null hypothesis could be rejected at the one per cent level of confidence was then compared with a similar testing of differences between repeaters and first offenders among boys.

As contrasted to boys, repeating among girls was linked to a narrow range of factors. The girls appeared to be more influenced by present relationships within the home and particularly by the mother-daughter relationship. The largest clusters of items associated with female repeating related to friction with parents and with school. The only direct socio-economic factor positively linked to repeating was size of home. In this case it was found that the relatively few girls living in units with two or more rooms per person were likely to be repeaters, if they came to police attention at all.

These findings raised a question as to the extent to which there might be a distinctively feminine pattern not only of offenses but

of associated variables. To test this out it was decided to undertake a comparison between all girl repeaters and all boy repeaters who came to the attention of the Youth Bureau of the Detroit Police Department in 1952.¹ Although there are obvious limitations upon data collected by the police, there are compensating advantages. One very important fact is that records made at this point in the "processing" of a "delinquent" eliminate the biasing influence of the selective factors which determine official action at the Juvenile Court level. For instance, in handling the repeaters with whom we are concerned, the police invoked Court action for 58.5 per cent of the boys but only 35.2 per cent of the girls. By contrast, dispositions involving referral to various social agencies were made for 13.8 per cent of the boys and 30.0 per cent of the girls. Clearly, a comparison of girls with boys declared officially delinquent by juvenile courts measures the selective factors as much as any basic sex differences between the two groups. Even the comparison at the police level is affected by selective factors related to attitudes resulting in complaints.

The procedure used in the present study was simple. The records on file for 1952 were examined and those for youngsters who had previous offenses were segregated. In the processing of each complaint, the Youth Bureau of the Detroit Police Department as a matter of routine undertakes an investigation which includes interviews with youngsters and parents, and visits to the homes. The information gathered is recorded on a "history sheet," and then transferred to IBM cards. For 1952, these permitted sixty-one tabulations related to the young person, his home, school, neighborhood, leisure, and offenses.

For 1952, there were totals of one thousand, four hundred and eighty-three boy repeaters and four hundred and twenty-seven girl repeaters. When the chi-square test was applied to the sixty-one tabulations, it was found that for fifty-one the null hypothesis could be rejected at the one per cent level of confidence, and for an additional four, at the two per cent level. Clearly, there are substantial differences between boy and girl repeaters. These are summarized in Table I.

¹ For aid in undertaking this study and making police records available, the author wishes to express appreciation to the following members of the Youth Bureau: Senior Inspector Sanford Shoults, Inspector Ralph Baker, Lieutenant Francis Davey, Sergeant Elizabeth Grindle, and Sergeant Mervin Roberts.

TABLE I—ITEMS LINKED TO SEX AMONG JUVENILE REPEATERS

Item	Categories High for Girls	Categories High for Boys	P<
Years in Detroit	0; 1	Life	0.01
Size of dwelling unit	More than 1.5 rooms per person	Less than 1 room per person	0.01
Type of dwelling unit	Rooming house; apartment	Single family homes	0.01
Condition of dwelling	Good	Average; needs minor repairs	0.01
Rating of neighborhood	Above average; below average	Average	0.01
Type of land usage	Mixed residential and commercial	Residential	0.01
Parent employed	Mother; neither	Father only	0.02
Income of parents	Regular low	Adequate	0.01
Recreational equipment of youngster	Less than playmates	Same as playmates	0.01
Distance from nearest recreational facility	20 blocks or more	5 blocks or less	0.01
Condition of interior of dwelling	Meticulous	Average; slovenly	0.01
Step-parents	Yes	—	0.01
Marital status of parents	Divorced	Home intact	0.01
Family quarreling	Frequent	Little or none	0.01
Youngster living with:	Stepfather and mother; boarding home	Both own parents	0.01
Father at home	Nights	Never	0.01
Mother at home	Never	All the time	0.01
Parents' attitude toward youngster	Punitive; rejecting; over-protective	Forgiving	0.01
Youngster's expressed attitude toward parents	Likes mother and doesn't mention father; dislikes mother and doesn't mention father; dislikes father and doesn't mention mother	Likes both parents	0.01
Youngster's expressed attitude toward home	Dislikes	Likes	0.01
Attitude toward mother	Dislikes	Likes	0.01

TABLE I—*Cont.*

Item	Categories High for Girls	Categories High for Boys	P<
Attitude toward father	Not mentioned	Likes	0.01
Attitude toward step-mother	Dislikes	Likes	0.01
Participation of parents in youngster's activities	Regular; occasional	None	0.01
Can youngster drive family car?	No	Yes	0.01
Does youngster complain of being "picked on"?	Yes, by mother; by father	No	0.01
Do parents approve of dating?	No	Yes	0.01
Alcoholism in home	Yes, leading to family discord	None; Yes, leading to abuse of youngster	0.01
Chores around home	Some	Few; none	0.01
Parents' attitudes toward police	Indifferent; antagonistic; trying to alibi selves	Coöperative	0.01
Church attendance	Regular; less than once a month	Less than once a week; never	0.02
Attitude of youngster to adult neighbors	Indifferent	Friendly	0.01
Attitude toward school	Dislikes	Likes	0.01
School grades	Good	—	0.01
Grade in school	Ninth	—	0.01
Attitude to teachers	No comment	Special class	0.01
Relations with classmates	Quarrels	Friendly	0.01
Member of:	School clubs	Friendly	0.01
Peer group relations	"Lone wolf"	Boy Scouts; YMCA Member of regular play group; gang	0.01
Nature of peer group activities	Orderly	Theft	0.01
Favorite sport	Dancing; not stated	Games	0.01
Favorite entertainment	Dancing; not stated	Shows	0.01
Physical development	Large for age	Small for age	0.02
Sexual development	Mature	—	0.01
Appearance	Neat and clean	—	0.01
"Mentality"	Above average	—	0.02

TABLE I.—*Cont.*

Item	Categories High for Girls	Categories High for Boys	P <
Age	13; 14; 15	12 or under; 16	0.01
"Attitude" toward officer	"Anti-social"		0.01
"Type"	Quarrelsome	Mischievous	0.01
Paid employment	None	Part-time; full-time	0.01
Type of employment	—	Labor; newspaper	0.01
Disposition of earnings	—	Part to parents	0.01
Offences (At age 12 or under)	Incorrigible; truant from home	Breaking and entering; malicious destruction of property	0.01
Offenses (At age 13 or 14)	Sex offenses; incorrigible; truant from home	Breaking and entering; larceny; auto theft	0.01
Offenses (At age 15 or 16)	Sex offenses; disorderly conduct; incorrigible; truant from home	Assaults; breaking and entering; larceny; and auto theft	
Police disposition of case	Released; referred to social agency	Complaint filed; referred to probation officer	0.01

It will aid interpretation of that table to recognize some characteristics of the chi-square test when applied to large groups. First, the chi-square calculation is based on numbers, rather than on intensity. It is a comparison of groups, not of individual qualities. Thus, the table cannot be taken to show that the average girl repeater hates her mother and the average boy loves his. Rather, among girls a larger number expressed dislike. More important, the chi-square is raised by items where there is a difference, even though only a small minority is involved. For instance, it would be a mistake to conclude that girl repeaters are transients and boys, native Detroiters. On this item, to be specific, 3.3 per cent of the girl repeaters but only 0.8 per cent of the boys had been living in the city for less than a year. Altogether, 82.0 per cent of the boys and 73.8 per cent of the girls had lived in Detroit all their lives. This attribute of the chi-square also permits measurement of the tendency for a group to show variability. Thus, on ratings of neighborhood quality, the girls showed greater dispersion; more

of them came from areas classed as "above average" and "below average." With the resultant cautions in mind, let us now compare the girls as a group with the boys as a group.

On socio-economic indices, the girls showed a wider range. More of them came from transient families living under conditions of relative poverty. However, the larger population also included a minority coming from economically advantaged homes, as attested by dwellings with 1.5 or more rooms per person. Although there was a similar minority among boys, for girl repeaters this minority was larger in terms of percentages. (For girls, eleven per cent came from homes having 2, 2.5, or 3 rooms per person; for boys the percentage was 6.9.)

Disruption of the home affected proportionately more girl repeaters than boy repeaters. For instance, more girls had divorced parents, lived with a step-parent, or were in boarding homes. More girls had working mothers, or, if both parents were home, reported severe quarreling.

Open friction at home was more prevalent for girls. This was evidenced by the fact that at all ages girls were frequently charged with "incorrigibility," with the charge for this offense originating with parents or step-parents. Another indication is that among youngsters having step-mothers, 18.4 per cent of the boys and 59.1 per cent of the girls expressed dislike. The attitude toward their own mothers showed the same trend, but to a lesser degree. Interestingly, attitude toward father or step-father was not as clear-cut. Girls, more often than boys, complained of being picked on. When interviewed by the police, more of the girls' parents fell into the minorities displaying hostile, rejecting or over-protective attitudes.

On one index of family functioning, the girls showed a favorable picture; more came from homes where parents participated regularly in their child's activities. Also, girls tended to have more chores to hold them in the house. This "closeness" may have been a disadvantage if the time spent in parent-child activity was typified by tension.

More of the girl repeaters showed lack of friendliness with other adults. More were indifferent to adult neighbors and disliked school. Interestingly, despite this, more of the girls earned good grades in school.

Relationship with peers showed interesting differences. The boy

repeaters were more often reported to be active in games, members of gangs, and getting in trouble with their gangs. More of the girls quarreled with their peers, and were "lone wolves." They belonged in greater number to sedentary school clubs, as opposed to the active Boy Scout and YMCA programs which attracted the boys. The livelier interest in sex was indicated by the high rating many girls gave to dancing. (It was checked as a favorite sport or hobby by 27.9 per cent of the girls but only 2.3 per cent of the boys.)

Aside from sex activity, more of the girls followed conventional behavior patterns as attested by regular church attendance, good school grades, membership in orderly peer groups, performing chores at home, and presenting a neat and clean appearance. In terms of attitudes, as these were rated by the police, however, more girl repeaters presented a picture of being quarrelsome, antagonistic, and resistant.

Another difference between the sexes results from differences in employment possibilities. Fewer girls obtain paid work.

This study verified the often-reported contrast in offense patterns. The boys tended to concentrate in the theft groupings and, among the older boys, to be charged more often with assaults. The younger girls were charged frequently with incorrigibility and running away from home. Above the age of thirteen, sex charges were often reported. As previously reported, police dispositions for boys made greater use of official court facilities; for girls, of referral to social agencies or release to parents.

CASE STUDIES

The distinctive quality of the factors surrounding girls' delinquency tended to be verified by case studies, of which three will be reported.

Gladys Nelson² had a record of five "arrests," two of them on complaint of her mother. The first, at sixteen, was an aftermath of a drinking party. At first she denied the charge of committing an immoral act, but when medical examination revealed a recently ruptured hymen, she confessed, saying she had not wanted to be called chicken. Later offenses involved drinking, beating up a younger girl in a fight over a boy friend, and being found in a stolen car.

² To protect the privacy of individuals, all names and other identifying data are fictionalized.

The Nelson home is torn by parental dissension. Mr. Nelson drinks heavily, quarrels with his wife, and picks on Gladys. Although occupying the same domicile, the parents are technically separated. Mrs. Nelson has started divorce proceedings.

Mrs. Nelson uses Gladys as a maid, and complains the girl does not do her share of the household duties. Although the mother makes much of wanting to understand the girl, she nevertheless is trying to get her into a boarding home.

The Nelsons live in a five-room flat in a stable, middle-class, all-white neighborhood. Besides the parents and Gladys, there is a younger brother. During interviews, Gladys gives the impression of a sophisticated young lady, who hates her father.

The second case, illustrative of "incorrigible" charges, concerns Helen Olson, fourteen, who was picked up three times on complaint of her mother. Mrs. Olson had a long list of misdemeanors to report: Helen kept late hours, truanted from school, even when her mother brought her to the building. Mrs. Olson also accused her daughter of immorality, which the girl denied.

The Olson family consists of the mother, daughter, and the mother's boy friend, all of whom occupy a two-room apartment in a working class district. The mother divorced Helen's father, remarried, and divorced the step-father.

Although Helen gets good grades in school, she does not want to attend. The family has moved so often that Helen is behind in her school work. When interviewed by the police, she cried and begged for another chance. However, three weeks later she was again truanting. She and a girl friend "walked around" during school hours. Helen said she was afraid of losing this friend, and the two had agreed to skip school every day.

The mother tends to treat Helen as though the girl is a nuisance. Mrs. Olson seems to be preoccupied with her own problems. After lodging complaints against her daughter, she was nervous and belligerent with the police.

Rarely would a similar situation be presented in the case of a boy. The truancy charge did not originate with the school. No other of Helen's actions would have drawn a police charge. Certainly, one never hears of a man living out of wedlock and asking the police to take action against a son for allegedly engaging in sexual intercourse.

A rather different picture is presented by Ida Phillips, fifteen,

whose record included three arrests for larceny and one escape from a detention home. In this case the family was intact. Mr. and Mrs. Phillips, with their three children, occupied a seven-room flat in an average, solidly Negro part of the city.

Interviews with mother and daughter tended to bring out a sham surface of affection. Mrs. Phillips at times took the tack of accusing the police of trying to "get" the girl. Then, later, she wanted to have the girl placed in a private institution for wayward girls.

In the day-to-day operation of the home, Ida was left a great deal to herself, Mrs. Phillips led an active social life centering in club activities. The father was a wishy-washy nonentity whom the mother kept out of the case.

In many respects, the pattern followed by Ida Phillips could be duplicated by many boy repeaters.

DISCUSSION

Although there is undoubtedly much overlapping between the sexes on factors found in repeaters, there is also evidence that a distinctively feminine pattern occurs in enough cases to affect statistical indices. This pattern tends to be brought to a focus in the fact that a parent, usually the mother, is the complainant. This act often results from a long history of tension. Whereas boys tend to escape their homes by engaging actively in sports and gang activities, some of the girl repeaters get caught up in squabbles. As a consequence a number become police cases by reason of acts which would be ignored in the case of boys.

There has often been an assumption that the comparatively large percentage of girls charged with sex-linked activities and with incorrigibility was due to the fact that our culture reflects a double standard of morality. While this may account for some of the difference, this study would suggest that the process whereby the charge is made deserves attention.

When compared with boy repeaters, many of the girls in our sample seemed to be more conforming to acceptable behavior standards. The outstanding segment where they suffered by comparison was their more open hostility to the adults in their lives. Yet, as illustrated in the cases of Gladys Nelson and Helen Olson, this hostility emerges as a product of a nasty, personal feud within the home.

In the light of the above, one may well wonder whether the ques-

tion to be studied in the case of at least some of these girls should not be phrased, "What factors lead parents to invoke legal machinery to settle quarrels with their daughters?" or "What leads a girl to clash with her mother so much that the police are called in?" rather than "What are the causes of female delinquency?"

REFERENCE

- 1) William W. Wattenberg and Frank Saunders, "Recidivism among Girls," *Journal of Abnormal and Social Psychology*, 50: 405-406, 1955.

THE GENERAL ABILITY TO JUDGE SOCIO-METRIC STATUS: ELEMENTARY STUDENT TEACHERS' SOCIOMETRIC PERCEPTIONS OF CLASSMATES AND PUPILS¹

NORMAN E. GRONLUND

College of Education, University of Illinois

It has been definitely established that individuals can judge the sociometric status of group members with better than chance success, and that there are large variations among individuals in their ability to make such judgments (1, 5, 7, 8, 9). These two conclusions seem to lead to the question raised in the present study. Is there a general ability to judge the sociometric status of group members? If it is established that this ability is general, it will then be possible to determine the personality correlates of this ability, the extent to which it can be improved by training, and how this ability is related to other types of social perception.

In addition to its theoretical significance, it is of practical import to know whether the ability to judge sociometric status is general. The relationship an individual establishes with his peer group is related to his ability to judge accurately the sociometric status of the group's members (7). Other investigators (4) have also found a relationship between the accuracy of teachers' social perceptions and the effectiveness of their relationships with students. Thus, it appears that an individual who is a good judge of others will find it easier to make social adjustments. If the ability to perceive sociometric status is general, then some basis will be provided for predicting the same individual's adjustments to different groups. This would be especially valuable in the case of teachers, who must make adjustments to their colleagues (peer group) as well as to their pupils (non-peer group). With the first group it is a matter of good staff relationships that contribute to the smooth functioning of the school. With the latter group it is not only the development of good personal relationships with pupils, but also the ability to assist pupils with their social adjustments. The effectiveness of

¹ This research was supported, in part, by a grant from the Bureau of Educational Research, College of Education, University of Illinois.

teachers in this regard is partly dependent upon their ability to perceive accurately the relative sociometric status of their pupils.

The specific problem under investigation is that of determining whether elementary student teachers who are good judges of the sociometric status of the members of their own *peer group* (classmates) are also good judges of the sociometric status of the members of a *non-peer group* (pupils they are teaching). A significant relationship between the accuracy of their judgments in these two situations would indicate the presence of a general ability to judge sociometric status.

METHOD OF INVESTIGATION

Population. The population consisted of one hundred and seventy-six student teachers and seven hundred and seventy-six pupils. The student teachers were enrolled in fourteen sections of the course, *Technique of Teaching in the Elementary School*, and were concurrently doing their student teaching at the elementary level. The fourteen sections ranged in size from eleven to fifteen with an average of approximately thirteen students per section. The seven hundred and seventy-six pupils were in the third, fourth, fifth, and sixth grades of the Champaign-Urbana Public Schools.² There were thirty classroom groups represented. Class size ranged from seven to thirty-six with an average of twenty-six pupils per class.

Procedure. A sociometric test was administered to the student teachers in all fourteen sections of the course, *Technique of Teaching in the Elementary School*, three months after the beginning of the course. Since the class procedure consisted mainly of informal discussion, it was felt that the students were well enough acquainted with each other to make valid sociometric choices. The sociometric form consisted of an alphabetical list of the students in each section. On this form each student was asked to rank the students in his section in the order in which he would prefer them as teaching companions. This ranking was done from the extremes of the distribution toward the center. Thus, each student indicated the classmate he preferred *most* as a teaching companion, then the

² Grateful acknowledgment is extended to Dr. Fred Barnes and the supervising teachers of the Elementary Education Division, College of Education, University of Illinois; to Miss Margaret Henderson, Director of Elementary Education, Champaign Public Schools; to Mr. Alvin King, Director of Elementary Education, Urbana Public Schools; and to the principals and teachers participating in the study, for their fine coöperation.

student he preferred *least* as a teaching companion, the one he preferred *next most*, *next least*, etc., until all of his classmates had been ranked. The relative sociometric status of the students in each section was obtained later by averaging the rank-order preferences of each individual.

After the sociometric form had been completed, the students in each section were asked to judge the relative acceptability of their classmates as teaching companions. This was done on another alphabetical list of the class members. Before making their judgments, it was explained to the students that a compilation of the sociometric preferences would indicate the relative acceptance of each student, by the group, as teaching companion. They were then asked to predict this rank order. Again the ranking was done from the extremes of the distribution toward the center. These procedures provided the data for determining the accuracy of the one hundred and seventy-six student teachers' *peer group judgments*.

A selected number of the student teachers was requested to make judgments also of the relative sociometric status of the pupils they were teaching. Originally, all of the students doing their practice teaching at the third-, fourth-, fifth-, and sixth-grade levels were to be selected for participation in this part of the study. Of the eighty-seven student teachers in this category, thirty-nine were eliminated because the schools in which they were teaching found it impossible to participate in the study. Another eighteen were eliminated because they had already seen the results of a sociometric test administered to their pupils. Thus, data on the accuracy of student teachers' *non-peer group judgments* were obtained for thirty of the original subjects. Before making their judgments the student teachers read the following statement:

The pupils in your class will be asked to name the five children with whom they would *most prefer to work*. The total number of choices each child receives will indicate his (her) acceptability to his (her) classmates as *work companion*. Some children will be highly accepted, receiving many choices; some will remain unchosen; the remainder will fall in between these extremes. Consequently, the children can be ranked in the order of their relative acceptability as work companion from the most accepted to the least accepted.

We are interested in determining the extent to which student teachers can judge the relative acceptability of each child to his (her) classmates as work companion. As you know, boys and girls at the later elementary level seldom choose members of the opposite sex as companions. Therefore, we are asking you to judge the boys and girls in your class separately.

Following this statement, each student teacher was requested to rank all of the boys in his class in the order in which he thought the boys would be accepted by their classmates as work companion. The same procedure was followed in obtaining judgments of the relative sociometric status of the girls in class. This ranking, like that in the peer groups, was done from the extremes of the distribution toward the center. Since the student teachers had devoted half of each day for approximately three months to their student teaching, it was felt that they knew their pupils well enough to make such judgments.

The sociometric test, described above, was administered to the pupils in each of the thirty public-school classrooms by the regular teacher. The use of five choices was prompted by the greater stability of sociometric results (10, 12) with this number of choices. The relative sociometric status of the pupils in each class was determined later by totaling the number of choices each pupil received. All choices were given a value of one regardless of level of choice, since it has been shown (10) that the arbitrary weighting of choices does not increase the stability of sociometric results.

ANALYSIS OF DATA AND RESULTS

Accuracy of Peer Group Judgments. The accuracy of each student teacher's judgments of his classmates was determined by correlating his prediction of the relative sociometric status of the class members with the results of the sociometric choosing. The resulting correlation coefficients represent the *accuracy scores for each student teacher's peer group judgments*. The means and standard deviations of these accuracy scores, for each of the fourteen sections, are presented in Table I. Although the mean accuracy scores range from 0.37 to 0.68, the *t*-test has shown *no* significant differences among the means of the various sections. The standard deviations for each section, however, indicate considerable variation in judgment accuracy among the individuals within each section. Since Bender and Hastorf (2) have suggested the possibility that such variations in accuracy may be accounted for by different degrees of "projection" on the part of the individual aided or hindered by the amount of similarity present between the individuals and the persons judged, it was decided to investigate the possible influence of this factor on the accuracy scores.

In order to do this, two scores, suggested by the work of Gage and

TABLE I—MEAN CORRELATIONAL SCORES AND STANDARD DEVIATIONS REPRESENTING THE ACCURACY OF ELEMENTARY STUDENT TEACHERS' JUDGMENTS OF THE SOCIOMETRIC STATUS OF THEIR CLASSMATES ($N = 176$)

Class Section	Fall, 1953			Spring, 1954			Fall, 1954		
	<i>N</i>	<i>M</i>	S.D.	<i>N</i>	<i>M</i>	S.D.	<i>N</i>	<i>M</i>	S.D.
A	12	0.40	0.16	13	0.51	0.14	12	0.54	0.17
B	11	0.64	0.20	13	0.53	0.21	12	0.37	0.18
C	11	0.50	0.17	14	0.59	0.18	13	0.41	0.25
D	12	0.68	0.18	13	0.57	0.14	15	0.66	0.16
E	12	0.43	0.22	13	0.64	0.21			

Exline (3), were determined for each individual. One was a similarity score, and the other an assumed similarity score. The *similarity score* was obtained by correlating each individual's sociometric preferences toward his classmates with the group members' relative sociometric status. Thus, the resulting correlation coefficients indicated the degree to which each individual's sociometric preferences were similar to the sociometric preferences of the group. The *assumed similarity score* was obtained by correlating each individual's sociometric preferences toward his classmates with his judgments of their sociometric status. Although this correlational score indicated the extent to which the individual's preferences toward others is related to his judgments of others, it cannot be assumed that it represents only "projection" on the part of the individual. If similarity scores are high, then high assumed similarity scores on the part of an individual might, to a large extent, represent insight into the attitudes and feelings of the group members. The present data do not make clear to what extent the assumed similarity scores are based on insight and to what extent they represent projection, but there is some evidence that both are probably present.

The interrelatedness of the similarity scores, the assumed similarity scores, and the accuracy scores, for each individual, may be seen in Figure 1. Since the three types of scores are correlational scores, the connecting lines represent the correlations between the three sets of responses.

The similarity scores, in the fourteen sections, ranged from a mean correlation of 0.36 to 0.63 with a median of 0.53. The assumed

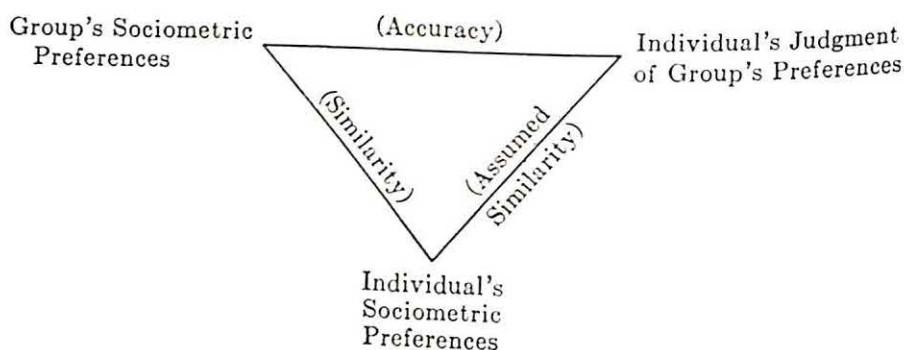


FIGURE 1

similarity scores ranged from a mean correlation of 0.58 to 0.74 with a median of 0.67. An inspection of the twenty-eight correlational scores in the fourteen sections revealed that in every section the mean assumed similarity score was greater than the mean similarity score. This indicates that individuals assumed greater similarity between their sociometric preferences and the preferences of the group than actually existed. This general tendency of individuals to overestimate the amount of similarity present would tend to indicate that "projection" was present in the assumed similarity scores. Since the correlational accuracy scores, the similarity scores, and the assumed similarity scores are interrelated through a variable common to each, this overestimation of similarity would tend to reduce the mean accuracy scores found in the fourteen sections. However, this "projection factor" may still inflate individual accuracy scores where similarity happened to be high.

In order to determine the possible influence of "projection" on individual accuracy scores, it was decided to compare the assumed similarity scores and the similarity scores for the two most accurate and the two least accurate judges in each of the fourteen sections. Before making this comparison, all of the scores were converted to *T*-scores. The twenty-eight *most accurate* judges had a mean accuracy score of 62 (*T*-scores), a mean assumed similarity score of 55, and a mean similarity score of 55. The twenty-eight *least accurate* judges had a mean accuracy score of 34, a mean assumed similarity score of 46, and a mean similarity score of 46. The *t*-test has shown that the most accurate group had significantly higher (0.01 level) assumed similarity and similarity scores. Thus, it is possible that the accuracy scores of the most accurate group were inflated to some extent by a combination of "projection" and higher similarity scores. However, as pointed out previously, it is also possible that the most accurate judges were more accurate in perceiving the

amount of similarity present and their assumed similarity scores, thus, represent insight into the feelings of the group rather than "projection" of their own sociometric preferences. Nevertheless, the large difference in accuracy scores between the most accurate and the least accurate groups could not be accounted for entirely by "projection," even if this were known to be a dominant factor in the assumed similarity scores. It should be pointed out, however, that the possible contamination of these accuracy scores by a "projection factor" would, in all probability, lower the amount of generality found in the accuracy of the judgments over peer groups and non-peer groups.

Accuracy of Non-peer Group Judgments. The accuracy of each of the thirty student teachers' judgments of their pupils was determined by correlating their predictions of the pupils' relative sociometric status with the results of the sociometric test. Since boys and girls were judged separately, there were two accuracy scores for each teacher. Their accuracy in judging the sociometric status of boys was represented by a mean correlational score of 0.59 with a standard deviation of 0.19. Their correlational accuracy scores for girls had a mean of 0.61 and a standard deviation of 0.20. The lack of a significant difference in the accuracy of student teachers' judgments of the sociometric status of boys and girls verifies similar results found with experienced teachers (8).

The total accuracy of each student teacher's judgments was obtained by averaging his correlational accuracy scores for boys and girls. The method of Z-transformation was used in the averaging. The resulting correlation coefficients represent the *accuracy scores for each student teacher's non-peer group judgments*. These correlational accuracy scores had a mean of 0.60 and a standard deviation of 0.12. This mean accuracy score is similar to that found in several other investigations (5, 8, 9) concerned with experienced teachers' sociometric perceptions of their pupils.

Although studies (5, 8) have reported no relationship between class size and the accuracy of teachers' sociometric perceptions, it was decided to check the influence of this variable in the present study. The resulting correlation of 0.11 between class size and the student teachers' accuracy scores was not significantly different from zero, verifying the previous findings. A similar lack of relationship was found between the spread of pupils' sociometric status scores and the accuracy of the student teachers' judgments. The resulting correlation was 0.09. This finding is not in agreement with other investigations (6, 9) which reported a slight, but significant, positive relationship between the two variables. The lack of rela-

tionship in this study can probably be accounted for by the restricted variation of the standard deviations among classes. These scores had a mean of 3.3 and a standard deviation of 0.6, indicating that the spread of sociometric scores was fairly similar in all thirty classes of pupils.

Since two previous investigations (1, 11) have shown a slight increase in the accuracy of teachers' sociometric perceptions with older age groups, the influence of this variable on the student teachers' accuracy scores was also investigated. An analysis of the data revealed that the mean grade level of the six most accurate student teachers was 4.3; for the six student teachers with average accuracy it was 4.6; and for the six least accurate student teachers it was 4.3. Both the most and the least accurate student teachers had judged the pupils in four fourth-grade classes and two fifth-grade classes.

It is apparent that the accuracy of the student teachers' non-peer group judgments was not influenced by the number of pupils they had in class, by the spread of sociometric status scores among their pupils, nor by the grade level of the pupils they were teaching.

Generality of Judgment Accuracy. It will be recalled that there were thirty student teachers who made judgments of the sociometric status of their classmates (peer group judgments) as well as judgments of the sociometric status of their pupils (non-peer group judgments). Thus, the generality of their judgment accuracy could be determined by relating the accuracy of the peer group judgments to the accuracy of their non-peer group judgments. However, before correlating these two sets of accuracy scores, it was necessary to convert them to *T*-scores.

The student teachers' peer group judgments had a mean *T*-score of 50 with a standard deviation of 10.4. Although these accuracy scores were obtained from fourteen different peer groups, they indicate a normal distribution of accuracy scores. When these accuracy scores were correlated with the student teachers' accuracy scores for non-peer group judgments, a correlation of 0.49 was obtained. This coefficient was found to be significant beyond the one per cent level. Thus, *these data are consistent with the hypothesis that the ability to judge the sociometric status of others is general.*

Although the size of the correlation coefficient tends to indicate that specific abilities are also present to a large degree, it is suspected that uncontrolled variables reduced the amount of generality reported. Since the peer group judgments were obtained from

fourteen different college classes and the non-peer group judgments from thirty different groups of pupils, there is a strong likelihood that situational factors influenced the judgment accuracy scores. This probability is supported by a previous study (9) which has shown that some groups are more difficult to judge than others, due to variations in the complexity of the social structure. Had it been possible to have all of the student teachers judge the same peer group and the same non-peer group it is felt that the index of generality would have been higher. Nevertheless, this finding that there is a general ability to judge sociometric status indicates the desirability of further explorations in the area.

IMPLICATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The findings indicate that some student teachers are generally good at perceiving the sociometric status of group members while others are generally poor. Thus, this study should make a contribution to the future selection and training of teachers as well as add to the theory of sociometric perception. However, there are several questions which should be answered before these results will find much practical application.

The extent to which accuracy of sociometric perception is related to other forms of social perception should be explored. In the classroom, teachers are expected to judge the knowledge, attitudes, interests, and a host of other characteristics about their pupils. Although it is easier to obtain objective evidence in some of these areas than in others, a major share of teachers' daily responses are probably based on judgments of their pupils. The appropriateness of teachers' responses is determined in part by the accuracy of their perceptions. Thus, it seems important to determine whether there is a general ability to judge the characteristics of pupils.

A second area that needs exploration is that of the extent to which the ability to judge others is subject to training. Teacher training institutions now proceed on the assumption that teachers can be trained to understand their pupils. However, a previous study (12) has shown no relationship between accuracy of sociometric perception and amount of college training in education or psychology. Social-perception skills may be found to be unresponsive to training, or it may be discovered that a special type of training is needed. Nevertheless, further research should be directed toward an evaluation of the trainability of these skills.

In addition to the above studies, it would seem important to investigate the specific rôle that social perception plays in teaching effectiveness. It may turn out to be a basic condition for effective teaching or merely a facilitating factor. However, the practical import of the findings would tend to indicate that research in this area is urgently needed.

SUMMARY AND CONCLUSIONS

This study was concerned with the extent to which the judgment of sociometric status is a general ability. The accuracy of elementary student-teachers' judgments of the sociometric status of *peer group members* (classmates) was correlated with the accuracy of their judgments of the sociometric status of *non-peer group members* (pupils they were teaching). An analysis and comparison of the data revealed the following findings.

(1) The accuracy scores of one hundred and seventy-six student teachers' judgments of the sociometric status of their classmates (*peer group judgments*) ranged from a mean correlation of 0.37 to 0.68 with a median correlation of 0.54, in fourteen college classes. However, these differences in means among the groups were *not* significant, due to the large variations in judgment accuracy among individuals within each group. In all fourteen college classes the student teachers assumed greater similarity between their sociometric preferences and the sociometric preferences of the group than actually existed. This "projection factor" tended to lower their judgment accuracy.

(2) The accuracy scores of thirty student teachers' judgments of the sociometric status of their pupils (*non-peer group judgments*) attained a mean correlation of 0.60 with a standard deviation of 0.12. There was *no* difference in the accuracy with which student teachers judged the sociometric status of boys and girls. There was also *no* relationship between the student teachers' accuracy of judging pupils' sociometric status and the following variables: number of pupils in class; spread of sociometric status scores; and grade-level of the pupils judged.

(3) The *T*-scores representing accuracy of *peer group judgments*, for the thirty teachers participating in both phases of the study, had a mean of 50 and a standard deviation of 10.4, indicating a fairly normal distribution of scores. When these accuracy scores were correlated with the accuracy scores for *non-peer group judg-*

ments, a correlation coefficient of 0.49 was obtained. This coefficient was found to be significant beyond the one per cent level, indicating the presence of a general ability to judge the sociometric status of others.

A discussion of the results led to the suggestion that further investigation should be directed toward determining the relationship between sociometric perception and other forms of social perception, the trainability of social-perception skills, and the specific rôle that social perception plays in teaching effectiveness.

REFERENCES

- 1) D. P. Ausubel, H. M. Schiff and E. B. Gasser, "A Preliminary Study of Developmental Trends in Sociopathy: Accuracy of Perception of Own and Others' Sociometric Status," *Child Development*, 23: 111-128, 1952.
- 2) I. E. Bender and A. H. Hastorf, "On Measuring Generalized Empathic Ability (Social Sensitivity)," *Journal of Abnormal and Social Psychology*, 48: 503-506, 1953.
- 3) N. L. Gage and R. V. Exline, "Social Perception and Effectiveness in Discussion Groups," *Human Relations*, 6: 381-396, 1953.
- 4) N. L. Gage and G. Suci, "Social Perception and Teacher-Pupil Relationships," *Journal of Educational Psychology*, 42: 144-152, 1951.
- 5) N. L. Gage, G. S. Leavitt and G. C. Stone, "Social Perception in the Classroom," Champaign-Urbana, Bureau of Educational Research, University of Illinois, 1955. Report No. 5, Studies in the Generality and Behavioral Correlates of Social Perception (mimeographed).
- 6) N. E. Gronlund, "Generality of Teachers' Sociometric Perceptions: Relative Judgment Accuracy on Several Sociometric Criteria," *The Journal of Educational Psychology*, 47: 25-31, 1956.
- 7) N. E. Gronlund, "Sociometric Status and Sociometric Perception," *Sociometry*, 18: 92-98, 1955.
- 8) N. E. Gronlund, "The Accuracy of Teachers' Judgments Concerning the Sociometric Status of Sixth-Grade Pupils," *Sociometry Monographs* No. 25, p. 62, 1951.
- 9) N. E. Gronlund, "The Relative Ability of Home-Room Teachers and Special-Subject Teachers to Judge the Social Acceptability of Pre-Adolescent Pupils," *Journal of Educational Research*, 48: 381-391, 1955.
- 10) N. E. Gronlund, "The Relative Stability of Classroom Social Status with Unweighted and Weighted Sociometric Status Scores," *The Journal of Educational Psychology*, 46: 345-354, 1955.
- 11) J. L. Moreno, *Who Shall Survive?* p. 440. Washington, D. C., Beacon House, 1934.
- 12) W. I. Newstetter, M. J. Feldstein and T. M. Newcomb, *Group Adjustment: A Study in Experimental Sociology*, p. 154. Cleveland, School of Applied Social Sciences, Western Reserve University, 1938.

THE RELATION OF INTELLIGENCE AND PHYSICAL MATURITY TO SOCIAL STATUS OF CHILDREN¹

RICK F. HEBER²

Michigan State University

There have been a number of studies attempting to get at the relation of IQ to social status. Results have been very conflicting. Several studies have reported no significant relation. A large number of studies have found a significant but low positive relationship between the two, using the correlation technique.

It appears significant that most of the studies reporting an insignificant or at best, a low positive relationship, have used the correlation method, which of course assumes a rectilinear relationship. Now, if intelligence and social status are related, but not in a rectilinear manner, correlation might not detect it. If we divide the data into halves or thirds and compare means, we have a much better chance of discovering a relationship which is curvilinear.

In a relevant study, Bonney (1) obtained a low insignificant correlation between IQ and social status with second grade children. However when she then took the upper and lower quartiles on social status scores, she found only one person in the upper quartile with an IQ below 90 while there were sixteen in this classification in the lower quartile.

The author suggests that intelligence is important in regard to social status only to a point; that only in those individuals who show considerable deviation from the norm of the group in IQ, will there be a relation to social status.

The same questions arise in connection with research on the relation of physical maturity to social status in pre-adolescent children. Jones and Bayley (2) took two extreme groups of male pre-adolescents in physical maturity, using skeletal age as the criterion

¹ Special acknowledgment is due to Dr. Donald M. Johnson, Professor of Psychology, Michigan State University, for his help in the preparation of this paper. The author is also indebted to Dr. Charles DeLong, Assistant Professor of Education, Michigan State University, who made available for this study the records of the Child Development Laboratory, Michigan State University.

² Now at George Peabody College, Nashville, Tenn.

and compared them on reputation tests which were administered to the classes from which the sample was drawn. They found the advanced maturity group tended to be more popular while the low maturity group tended to be rejected.

The present study is designed to see whether individuals who deviate considerably from the group in either IQ or physical maturity also show a deviation in social status.

SAMPLE

This sample was made up of three classes of children, one class from each of the third, fourth, and fifth grades from a public school in a small suburban town. These children are both rural and urban, from predominately middle-income homes. Total number of children making up the study was ninety-seven, forty-eight girls and forty-nine boys.

PROCEDURE

Kuhlman-Anderson IQ scores were used for the intelligence measurements. Selecting some measure of physical maturity presented more difficulty. Jones and Bayley selected their high and low maturity groups using skeletal age as the criterion. They then found that the mean heights of these groups were far apart. It was decided to use height as the criterion of maturity in our study, bearing in mind the grossness of differentiation with use of such a single criterion. The Cunningham Social Distance Scale (S) was used to obtain a measure of social status for the individual. On this scale each pupil prints the name of each other pupil in the class in one of five categories assumed to be on a continuum from positive to negative. In computing the social status score of a child, the categories are weighted 5, 4, 3, 2, 1 from most positive (I would like him for a very close friend) to most negative (Wish he weren't in our room).

PRESENTATION OF RESULTS

Mean IQ's, heights, and social status scores are presented in Table I with the standard deviations.

Because of variations in height between boys and girls and between grades, *T* scores were computed for each group for height, IQ and social status. It was necessary to select some figure as a cut-off point to obtain a sample of children showing considerable deviation from the mean. Solely because of the necessity of obtaining

TABLE I—MEAN IQ, HEIGHT, AND SOCIAL STATUS

	Kuhlman-Anderson IQ's	Height in Inches		Social Status on Social Distance Scale
		Boys	Girls	
3rd Grade mean	102.57	51.20	51.62	3.55
SD	9.56	2.50	2.17	0.647
4th Grade mean	91.93	53.85	53.91	3.39
SD	11.07	2.49	1.89	0.660
5th Grade mean	94.20	55.99	57.58	3.73
SD	12.75	2.42	2.86	0.410

TABLE II—SOCIAL STATUS SCORES FOR HIGH, AVERAGE AND LOW IQ GROUPS

	High IQ				Average IQ				Low IQ			
	No.	K-A IQ	Height	Social status	No.	K-A IQ	Height	Social Status	No.	K-A IQ	Height	Social status
High maturity	4	64.2	64.0	60.5	10	50.7	66.1	51.8	3	37.4	64.4	42.3
Average maturity	11	63.0	50.0	59.7	36	51.2	51.0	52.2	10	36.2	49.6	33.5
Low maturity	4	62.9	36.7	50.5	11	50.7	37.3	54.4	8	33.6	37.1	39.1
Total	19	63.2	50.1	58.7	57	51.0	51.0	52.6	21	35.2	47.0	36.9

a fair-sized sample, a standard deviation of 0.84 was selected. There was no other basis for choosing this figure. In future studies with large samples it might be feasible to obtain finer breakdowns with more categories, thus permitting a more detailed analysis.

Thus with our distribution in *T* scores, all individuals with *T* scores in IQ or height of 58.4 or above were placed in the high IQ or advanced maturity category while those with a *T* score of 41.6 or below were placed in the low IQ or low maturity category. Those children with *T* scores between 58.4 and 41.6 were placed in the average category.

In Table II the social status scores in *T* scores are presented with breakdowns into high, average and low IQ categories and high, average and low maturity categories. The total mean social status scores for the high, average and low IQ groups are also presented

TABLE III—SOCIAL STATUS SCORES FOR HIGH, AVERAGE AND LOW MATURITY GROUPS

High Maturity				Average Maturity				Low Maturity			
No.	K-A IQ	Height	Social status	No.	K-A IQ	Height	Social status	No.	K-A IQ	Height	Social status
17	51.5	65.3	52.2	57	50.8	50.6	50.4	23	46.9	37.1	48.4

in Table II. The total mean social status scores for high, average and low maturity groups are presented in Table III.

None of the differences between mean social status scores for high, average, and low maturity groups are significant, either when taken as a total or when taken with the average IQ groups. In fact when the social status scores for the high, average and low maturity groups are compared in the average IQ category it will be seen that they are all extremely close to the mean of 50.

The total mean social status score for the high IQ group is 58.7, for the average IQ group, 52.6, and for the low IQ group, 36.9. All of these differences are significant beyond the 0.01 level by the *T* test.

DISCUSSION OF RESULTS

We found that there were no significant differences on social status between high, average and low maturity groups. This means one of two things: either our use of height as a measure of physical maturity was not a good one, or physical maturity is not related to social status at this grade level. Because of the rather close correspondence of height at the pre-adolescent level to other measures of physical maturity we might have used, we would expect to have obtained a difference in our distribution if the relationship here were a gross one. However further studies will have to be made using other criteria of maturity before a definite conclusion can be drawn.

In our IQ groups the differences on social status between our high and average groups and between our average and low groups were highly significant. In our sample then, the extreme deviates in IQ do show a deviation in social status. The high IQ group was 6.1 above the average group in social status, while the average IQ group is 15.7 points above the low IQ group. Thus the difference

between average and low is more than two and a half times as great as the difference between average and high. This would point toward the decremental effect on social status of a low IQ being greater than the incremental effect of a high IQ. Illustrating a more striking contrast, there is a difference of 21.8 points on the *T* scale for social status between high and low IQ groups.

Analyzing the data in another manner we find in the high IQ group only three children out of nineteen below the mean on social status. In the low IQ group there is only one child out of twenty-one above the mean on social status. Now in our average IQ group there were several children who went as high on social status as the highest of the high IQ group and several who went as low as the lowest of the low IQ group.

This may be the reason so many investigators have obtained a negligible correlation between IQ and social status. The relationship does not appear to be a rectilinear one. Rather it may prove to be an exponential function: that IQ is important only up to a point. We have evidence for this in the extremely low social status of our low IQ group.

CONCLUSIONS

(1) Physical maturity as measured by height shows no relation to social status in pre-adolescent children.

(2) Children who show considerable deviation in IQ from the group also deviate in social status in the same direction as the IQ deviation.

REFERENCES

- 1) M. E. Bonney, "A Study of Social Status on the Second Grade Level," *Journal of Genetic Psychology*, 60: 271-305, 1942.
- 2) Mary Cover Jones and Nancy Bayley, "Physical Maturing among Boys as Related to Behavior," *Journal of Educational Psychology*, 41: 129-148, 1950.
- 3) Ruth Cunningham, *Group Behavior of Boys and Girls*. New York, Bureau of Publications, Teachers College, Columbia University, 1951.

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN PERSONALITY AND VOCATIONAL INTEREST

WILLIAM R. MELTON, JR.

California Test Bureau, Los Angeles

A phase of the mushrooming guidance and testing movement that, according to all available evidence, has suffered in comparison with others is that which would *relate scientifically* the various factors of human development measured in today's guidance testing programs. Too many generalities characterize the thinking and too many broad assumptions have been made by counselors as well as test makers, at least in dealing with personality, interests, and aptitudes. Much work has been done by pioneers in the testing field to relate intelligence to achievement, and with well-known success. But, in spite of the fact that the measurement of interests is thirty-five years old, very little experimentation has been done to correlate interests with aptitudes (7, pp. 379-99) and even less to relate personality to interest. It is with this last aspect that this study has concerned itself.

It was the purpose of this investigation to determine what personality patterns, if any, are peculiar to those interested in certain occupations as evidenced by the results of administering to high school seniors standardized tests of vocational interest and personality. In effect, are those persons interested in becoming teachers, for example, high in social skills and community relations and perhaps low in nervous symptoms? Are those interested in entering the acting profession, for example, high in self-reliance and freedom from withdrawing tendencies and possibly low in family relations? Or, in terms of the layman, are actors "high-strung"? Are mechanics "easygoing"? etc. If definite patterns were to appear, the way would then be clear for other researchers to demonstrate the possible uses to which such knowledge could be put.

Much care and thought go into the making of modern standardized tests and an equal amount of care should be employed in their use. Counselors should be equipped, with the help of researchers, with the most successful and practical methods of approach once they have the results of a testing program in their hands. In spite

of how well-meaning their efforts may be, gross injustices to our children may be perpetrated unless those in charge of guiding them know how to relate significantly the various scraps of information they may learn about them. The present a priori basis for explaining such relationships as may exist can and should be replaced by a scientific approach.

Perhaps the well-chosen words of Cattell best describe the current status of the question:

"It is a sad illustration of the meager harvest accruing to pure science from comparatively heavy expenditures on applied science that, in spite of the enormous attention vouchsafed in the last forty years to the psychology of vocational guidance, we still have no figures even for the *means* of occupations with regard to the principal personality factors. . . ." (1, p. 418).

Comparatively little research is available which can be considered truly conclusive in relating personality to vocational interest. In previous studies the number of cases involved and the correlation coefficients obtained have not been high in most instances (seldom over ± 0.40); and often having to do with the masculinity and femininity of interests, results are usually couched in clinical terms unfamiliar to the average teacher or guidance worker. A good many investigators have been unable to conclude much more than that personality maladjustment is most commonly associated with arts and social service and that adjustment is highly correlated with interest in the mechanical field of endeavor. Most important to this study, very little work has been done with the instruments involved.

METHODS

The group involved in the study consisted of three hundred and twenty-four seniors taking a course in social studies at El Monte Union High School in El Monte, California. The median Language I.Q. for the group was 105; the median Non-Language I.Q., 102, indicating that the group was a very normal one as far as mental maturity was concerned. The group was almost entirely of the Caucasian race and divided evenly between males (50.5 per cent) and females (49.5 per cent). Ninety-one per cent of the population was from either the West or the Midwest. A student questionnaire was completed by each person, giving the author, in addition to the

above information, an idea of the socio-economic status and expressed interests of the group.

The following procedure was followed in administering standardized paper-and-pencil tests in personality and interests to the experimental group and in correlating results.

(1) First, students for whom scores on the California Test of Mental Maturity, Advanced, 1946 Edition (6) were available (so that Language and Non-Language I.Q.'s could be used to describe more adequately the various interest groups and predict their chances for success in their chosen profession) were given the California Test of Personality, Secondary, Form A (8).

(2) Then the group took the Occupational Interest Inventory, Intermediate, Form A (3).

(3) Using the seventieth percentile on the Occupational Interest Inventory as a criterion (cut-off point), which is selective enough to indicate true peaks of interest out of the six major occupational fields, data for the students of both sexes who met that criterion (composite norms were used) were selected for further testing and analysis.

(4) Then two hundred and twenty-nine students who met the above criterion were given the Vocational Interest Analyses (5) which corresponded to interests registering at the seventieth percentile or above on the Occupational Interest Inventory. Since most students had at least two peak interests, the basic data were nearly doubled. It should be noted here that the Vocational Interest Analyses are a new, six-fold analytical extension of the Occupational Interest Inventory.

(5) Of the four hundred and twenty Vocational Interest Analyses taken, seventy-nine of them were in the Personal-Social field; eighty-two, Natural; sixty-eight, Mechanical; ninety-three, Business; ninety-four, Arts; and only four, Sciences. Since each of the six Analyses is subdivided into six more specific occupational fields a criterion had to be set up which would indicate true interest in them on the part of the examinee as it was for these thirty-six subgroups that mean personality profiles were to be computed. Also correlations between interest scores and personality scores were to be computed on them. It was decided to use only those cases the scores of which were above the midpoint of the range as a criterion to select those who would appear to be truly interested in any of

the thirty-six specific types of activities on the Vocational Interest Analyses. A cut-off point of 22 was used rather than 20, since the standard error of measurement on nearly all of the Vocational Interest Analyses sub-groups lies between two and three score points, and a score of 22 or better is thus more likely to be genuinely above the midpoint of the range. Cases which did not meet this criterion were discarded from the study. In all, nine hundred and thirty-four sub-scores were available for further study at this point.

(6) Sufficient interest was shown in twenty-six of the thirty-six interest groups, and *N*'s varied from sixty-seven (Social Service) to eighteen (Bench Work and Bench Crafts) with an average of thirty-six. Results for the entire Sciences Analysis were omitted since an almost complete lack of interest was shown for any occupation involving scientific work by the group.

(7) The results of the above testing program were punched on Hollerith cards, sorted, and tabulated to form scatter diagrams preliminary to the hand computation of correlations, means, and standard errors of measurement.

(8) Scores of the twenty-six interest groups on the Vocational Interest Analyses were correlated by means of an adaptation of the tetrachoric method with scores on the twelve components, Self, Social, and Total Adjustment on the California Test of Personality. This involved $390 (26 \times 15)$ correlations in all, using the formula:

$$r_t = \cos \pi \frac{\sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

In each correlation a scatter diagram was formed first as for a Pearson *r*. Then, to obtain a true median split, the two class intervals in which the medians fell were ruled off, thus presenting a 3×3 scatter plot, in each cell of which the frequencies were totaled. The frequencies in the median intervals were then distributed proportionally to the four corner cells. At this point, the formula above was applied.

It should be noted here that the formula was developed by Pearson, himself, and he reported that the mean error of this method of estimating *r* is less than four per cent (4, p. 370). This has been experimentally verified by others and it is rare that an error greater than 0.05 is encountered, when care is taken to ensure median dichotomy on both axes. The smoothing procedure described above performs the additional function of avoiding the great effect that

chance fluctuation exerts when N is small, as it is in some of the correlations involved in this study.

As a test of significance, standard errors of measurement were computed for the three hundred and ninety correlations, using the formula: $SE_{r_{t_o}} = \frac{1.5708}{\sqrt{N}}$, which is derived from Pearson's formula

for computing the probable error of a tetrachoric r ($PE_{r_{t_o}} = \frac{0.6475 \pi}{2\sqrt{N}}$) when the dichotomic lines are at the medians in both arrays (4, p. 372). To determine the level of significance from these SE 's, the r 's were divided by SE_r and the results looked up in Fisher's table of the value of t at the various levels of significance (2, p. 330).

In the case of the smaller samples ($N < 60$), the formula $t = \frac{\sqrt{r^2(N-2)}}{1-r^2}$ was used and the result multiplied by $\frac{\pi}{2}$ to provide a more rigorous test (4, p. 154). For example, a minimum value of a tetrachoric r significant at the ten per cent level is given by $r =$

$\frac{\frac{\pi}{2} t_{10}}{\sqrt{N-2+t_{10}^2}}$ in which t_{10} (the value of t significant at the ten per cent level) was obtained by simply referring to Fisher's table.

(9) Median scores on the California Test of Personality were computed for the twenty-six interest groups and then translated into percentiles. These were plotted on profile sheets for the personality test. The middle fifty per cent of each group was also plotted on the profiles to show the scatter of the scores.

RESULTS

Among the results of the correlation study, the nine relationships shown in Table I were found to be significant at the five per cent level of confidence.

The fifteen correlations in Table II proved to be significant at the ten per cent level of confidence.

There were no correlations significant at the one per cent level of confidence; however, the author feels that the correlations in Tables I and II may very definitely indicate a trend.

The plotting of median personality profiles for the twenty-six interest groups brought to light some interesting results and they

TABLE I.—CORRELATIONS BETWEEN COMPONENTS ON THE CALIFORNIA TEST OF PERSONALITY AND INTEREST GROUPS ON THE VOCATIONAL INTEREST ANALYSES*

Component on the C. T. P.	Interest Group on the V. I. A.	N	Correlation Coefficient
School relations	Social service	67	0.44
School relations	Health and medical service	22	0.67
Social adjustment	Health and medical service	22	0.66
Total adjustment	Health and medical service	22	0.64
Community relations	Fish, game, and domestic fowl	39	0.51
Sense of personal freedom	Designing	29	0.63
Social skills	Bookkeeping and accounting	54	0.46
Self reliance	Management and control	36	0.52
Social standards	Drama and radio	51	0.44

* Significant at the 5 per cent level of confidence.

TABLE II.—CORRELATIONS BETWEEN COMPONENTS ON THE CALIFORNIA TEST OF PERSONALITY AND INTEREST GROUPS ON THE VOCATIONAL INTEREST ANALYSES*

Component on the C. T. P.	Interest Group on the V. I. A.	N	Correlation Coefficient
Self adjustment	Social service	67	0.33
Social standards	Social service	67	0.34
Total adjustment	Social service	67	0.32
Feelings of belonging	Health and medical service	22	0.60
Family relations	Fish, game, and domestic fowl	39	0.46
School relations	Fish, game, and domestic fowl	39	0.43
Self reliance	Maintenance and repairing	38	-0.49
School relations	Maintenance and repairing	38	-0.42
Total adjustment	Maintenance and repairing	38	-0.45
Feelings of belonging	Maintenance and repairing	38	-0.46
	Machine operation and tending	32	-0.39
Sense of personal freedom	Bookkeeping and accounting	54	0.50
Social standards	Management and control	36	0.51
Sense of personal worth	Art crafts	26	0.38
Social standards	Decorating and landscape	45	0.52
School relations	Literary activities	27	

* Significant at the 10 per cent level of confidence.

are described below. A tabulation of the best adjusted and most poorly adjusted groups on Self, Social, and Total Adjustment on the California Test of Personality was made and may be seen in Table III.

Table IV represents percentiles on the California Test of Personality which were so unusually low as to make them worthy of note.

The groups were most consistently high on Social Standards and Freedom from Anti-Social Tendencies.

Considering the five general interest fields under which the twenty-six sub-areas just analyzed are grouped, it was found that they were adjusted on Self, Social, and Total Adjustment on the California Test of Personality as shown in Table V.

As to the median intelligence quotients for the twenty-six groups, it is interesting to note that the general fields having the highest Language and Non-Language I.Q.'s were the two which were the

TABLE IV.—UNUSUALLY LOW PERCENTILES ON CALIFORNIA TEST OF PERSONALITY COMPONENTS EARNED BY CERTAIN INTEREST GROUPS ON VOCATIONAL INTEREST ANALYSES

Interest Group	C. T. P. Component	Percentile
Fish, game, and domestic fowl	Community relations	20
Maintenance and repairing	Social skills	20
Machine operation and tending	Social skills	20
Machine operation and tending	Community relations	20
Construction	Community relations	20
Shipping and distribution	Family relations	20
Management and control	Social skills	25
Gardening and greenhouse care	Freedom from withdrawing tendencies	25
Machine operation and tending	Freedom from withdrawing tendencies	25
Machine operation and tending	School relations	25

TABLE V.—ORDER OF ADJUSTMENT OF FIVE GENERAL INTEREST FIELDS ON SELF, SOCIAL, AND TOTAL ADJUSTMENT ON CALIFORNIA TEST OF PERSONALITY

Order of Adjustment	Self Adjustment	Social Adjustment	Total Adjustment
1	Personal-social	Personal-social	Personal-social
2	The arts	The arts	The arts
3	Natural	Mechanical	Natural
4	Mechanical	Business	Mechanical
5	Business	Natural	Business

best adjusted on the personality test: Personal-Social and The Arts.

The small amount of variance among the median I.Q.'s found in this study (which, incidentally were corrected to the 1951 norms), and the very fact that they are medians and do not disclose the extremes, makes prediction (especially on the basis of the small *N*'s), even with what we now know of intelligence requirements for success in given vocations, too hazardous a process at this stage.

CONCLUSIONS

On the basis of the foregoing results, the following conclusions as to the relationship between personality and vocational interest seem reasonable.

(1) Though high in a few but slight in a majority of instances (which is in line with previous research in this field), there are definitely measurable relationships between personality and vocational interest.

(2) Consequently, it would seem good psychological procedure to include the use of personality inventories as a must in guidance testing programs, and at as early a stage in an individual's career as is practicable, a view which some of today's counselors do not yet hold. Obviously, the earlier undesirable personality patterns are discovered and diagnosed, the sooner will guidance people be able to provide the therapy necessary for proper adjustment of those who evidence an interest in entering a certain profession. Certainly, in the face of the evidence, guidance personnel can no longer ignore the necessity of investigating further the relationship between personality and vocational interest to replace current *a priori* bases for prediction in this area.

(3) In view of the fact that results in this study indicated that those individuals interested in the general fields of Personal-Social and The Arts were the best adjusted as to personality and those interested in the Mechanical field the most poorly adjusted (which is in more or less direct contrast to the findings of previous studies), it would seem wise to examine more closely the findings of studies of this nature as well as intensifying research in the field to provide conclusive evidence as to the direction in which the truth lies.

(4) For the individuals involved in this study at least and possibly for other individuals, the California Test of Personality is especially effective as an indicator of personality trends or patterns

for those interested in Social Service, Health and Medical Service, Fish, Game, and Domestic Fowl, and Maintenance and Repairing, since these groups involved the greatest number of statistically significant correlations.

(5) Although it is commendable as an ultimate aim, on the basis of the results obtained in the present study it can only be concluded that, although the temptation is great, it would be too hazardous a course to use median personality profiles for the twenty-six interest groups as "templates" to judge future individuals expressing an interest in any of these twenty-six sub-areas of vocational activity. Counselors, themselves, however, can aid researchers in verifying such profiles in no small degree by accumulating evidence that those who express interest in certain occupations have certain personality patterns which are peculiar to all the others who profess interest in those occupations (if such is the case, as it appears to be). Thereby small N 's will become large N 's, skewed distributions will become relatively normal in comparison, and low correlations may in time become high ones, thus providing overwhelming evidence of relationships.

RECOMMENDATIONS

Since studies such as this involve certain necessary limitations, recommendations for further study become more or less self-evident. The more potentially fruitful of these recommendations in the author's opinion are listed here.

(1) Studies similar in nature to this one should be made with groups much greater in number. The number of cases should be extended to include students from all over the country, of all races, and socio-economic and intelligence levels. In addition to gaining for the researcher a more representative sampling of the nation's youth, increased numbers will enable him to analyze the personality of individuals interested in all of the thirty-six vocational subgroups, not merely twenty-six as in this study. Too, the age range might well be extended to include junior high school students, on the one hand, and adults (graduates) on the other; this, in spite of the present evidence of stability of interests. Obviously, the value of any statistical technique is enhanced by larger numbers of cases.

(2) Studies could be made to great profit with other standardized measuring devices than the ones involved in this and previous studies. By simply consulting yearbooks of test materials, the student

of educational psychology will see that there remain many tests which may be used in studies relating personality to vocational interest; and, although the task may appear gargantuan, it remains for someone to correlate results of analogous studies (but using different testing instruments) and emerge with the invaluable "general" pattern that could be used by guidance workers and counselors in guiding tomorrow's youth.

(3) As has been pointed out by other investigators and will be emphasized here, the relationship of personality not only to interests but to aptitudes and success on the job is needed and may well prove a fruitful field for future study. Since interest does not necessarily insure success and aptitude is not necessarily a correlate of interest, investigation of the one must be enforced by investigation of the others. The resulting continuum, were such research completed, would undoubtedly be invaluable to the guidance counselor, who could scientifically counsel an individual no matter whether his vocational strengths and weaknesses be in the interest, aptitude, or success-on-the job stage.

(4) Future research in this field should take into account much more than did this study the environmental (or hereditary) influences on vocational choices. The extent to which such things as transfer of interests from parent to child take place should be considered so that groups under consideration may be more adequately described. To insist that all interests are a function of personality unequivocally would be as big a mistake as to ignore its rôle in interests. The guidance worker is also obviously interested in such questions as: Is the student taking course work commensurate with his interests or is he enrolled in classes because his parents wish it? Does he feel that, in light of circumstances, he will be able to follow the career in which he is most interested? These problems have been considered independently by others, but are posed here to suggest that future investigators of the relationship between personality and vocational interest describe better their populations and the incidental influences on vocational choice.

(5) Last, but perhaps most important, it is recommended that the cause-effect relationship (which was beyond the scope of this study) be examined exhaustively through factor analysis so that hypotheses can be supplanted by facts and by action on the part of guidance people to improve or mould personality, on the one hand, and provide opportunities to work at jobs that will constitute ther-

apy on the other. Little, if any, evidence along this line is available at the present time to those charged with the responsibility of successfully guiding our children in the choice of their life's work. It is entirely possible that he who chooses to dedicate himself to such research may well earn the gratitude of countless personnel managers for reducing employee turnover to a minimum.

REFERENCES

- 1) Raymond B. Cattell, *Personality*. New York, McGraw-Hill, 1950.
- 2) Allen L. Edwards, *Statistical Analysis*. New York, Rinehart and Company, 1946; citing R. A. Fisher, *Statistical Methods for Research Workers*, p. 209, Edinburgh, Oliver and Boyd, Ltd., 1936; and George W. Snedecor, *Statistical Methods*, p. 149, Ames, Iowa, Collegiate Press, 1946.
- 3) Edwin A. Lee and Louis P. Thorpe, *Occupational Interest Inventory, Intermediate, Form A*. Los Angeles, California Test Bureau, 1944.
- 4) C. C. Peters and W. R. Van Voorhis, *Statistical Procedures and Their Mathematical Bases*. New York, McGraw-Hill, 1940; citing Karl Pearson, "On the Correlation of Characters Not Quantitatively Measurable," *Transactions of the Royal Society, Series A*, 195: 1-47; and R. A. Fisher, "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample," *Metron*, 1: 9, 1921.
- 5) Edward C. Roeber and Gerald G. Prideaux in collaboration with Edwin A. Lee and Louis P. Thorpe, *Vocational Interest Analyses*. Los Angeles, California Test Bureau, 1951.
- 6) Elizabeth T. Sullivan, Willis W. Clark, and Ernest W. Tiegs, *California Test of Mental Maturity, Advanced*, 1946 edition. Los Angeles, California Test Bureau, 1946.
- 7) Donald E. Super, *Appraising Vocational Fitness*. New York, Harper, 1949.
- 8) Ernest W. Tiegs, Willis W. Clark, and Louis P. Thorpe, *California Test of Personality, Secondary, Form A*. Los Angeles, California Test Bureau, 1942.

EDUCATIONAL THEORY AND THE PSYCHOLOGY OF LEARNING¹

G. T. BUSWELL

University of California at Berkeley

As background for this paper I should like to quote an illustration given by Guthrie (*1*) on the first page of this chapter in the National Society Yearbook on "The Psychology of Learning." It consists of a description of a day in school as told by a boy in the elementary grades. His account is as follows:

"We have a man teacher. He doesn't get mad much. We talk a lot. If somebody throws a pen or stuff like that, he puts you out in the hall. We have 20 words for spelling. We write them on Monday and Wednesday and Friday, but it only counts on Wednesday and Friday. Then we have social science. It's really history, but he likes fancy names for things. Then we study geography and then mathematics—really arithmetic, you know. He gives us some fancy doo-jigs to work with. About two boys get their seats changed if they've been poking somebody. Then the teacher gives us exams and stuff and we have recess—ten minutes—not much time for doing things. We mostly play marbles. Then the bell rings and we go in and the teacher talks a lot about notices. Then the music class. The music teacher talks a lot. We don't sing much. Sometimes we sing a review song, but mostly the teacher talks. Then we have reading. The teacher is nuts about ancient literature. Then we check our books at the book table. They have about 200 books. Some of them are good and some are corny. We have shop, too; make things. There's a jig saw and a lathe and a sander and a printing press. We use all those things. I made a record case and a broom holder and a box."

From the point of view of a boy, school consists of a summation of days and years of experiences such as those described. Certainly his account seems trivial and unimpressive as a sample of this multibillion-dollar enterprise we call Education. From the point of view of the psychologist and the educator, the behaviors described by the boy are likely to be evaluated in terms of the pattern of changes they produce in the boy as a human organism, in terms of the particular knowledge and skills that are developed, and in terms of the general intellectual and emotional habits that result. The educator will be concerned with the question of whether such

¹ Presidential address read at the meeting of the Educational Psychology Division of the American Psychological Association at the Annual Convention in San Francisco, September 10, 1955.

behaviors as those described by the boy will produce the kind of outcomes expected by society, and, if so, with the theoretical explanation of such outcomes. The psychologist will be concerned with a theory of learning that will explain the changes produced in the boy, although he will probably be less concerned about what outcomes are expected by society. The main concern of this paper is how a theory of education as expressed by the educator is related to a theory of learning as formulated by the psychologist.

In general, theories of education have been formulated quite independently of theories of learning and, likewise, learning theories have been set up by psychologists who are often quite innocent of an intimate knowledge of what goes on within schools. One of the main concerns of the Educational Psychology Division of the American Psychological Association, so it seems to me, should be to bring together the learning theory of the psychologists and the educational theory of the school.

At one of the section meetings of the American Association for the Advancement of Science, a competent educational psychologist made the statement that learning theory had no value for teachers in the classroom. Yet, since teaching implies learning, how can it be that a serious study of the theory of how one learns can have no value for the teacher? One is tempted to explain this opinion as due to the failure of the theorist to illustrate and to verify his theory in terms of the kinds of learning that are carried on in schools. For example, the classical experiment in conditioning is simple and clear in showing how substitute stimuli will produce the flow of saliva in the dog. One can illustrate the process of conditioning quite easily with young children by stimulating a fear response of running away from an assortment of stimuli. Here we have a constant, original response of secreting saliva, or of running, as a result of a variety of stimulations. However, the school is not overly interested in original responses. How shall the teacher use the process of conditioning to elicit all the varied kinds of responses that are the desired outcomes of education? When close to the level of native response, the conditioning theory is easily illustrated. But how shall we tell the algebra teacher to use it in teaching a ninth-grade pupil to square the binomial " $x + y$ "? Guthrie, in his text on the "Psychology of Learning" (2) asks the question, "What does the principle of conditioning mean in the form of practical advice?" He answers with an illustration on the

same page as follows: "If we wish to teach a dog to come when he is called, our method will be to get him to come by hook or crook. There are no rules for this except what we know of dogs in general." Is the teacher of algebra to be offered nothing better than a "by hook or crook" method? Or, is an unvarying production of a correct response all that is expected from the study of algebra? Further on, Guthrie states "... the student officer is cautioned never to give a command that he is not confident will be obeyed. If the command is followed by acts other than those commanded, the command becomes merely a cue for disobedience and the officer loses his authority." Is the teacher then never to ask a question until he is confident it will be answered correctly? And by what occult power will he know when the time is ripe? One may at least have some sympathy for the teacher who fails to find learning theory helpful. But in all truth, there is an obligation on the part of someone to make it helpful. It makes no sense when a serious study of learning and a serious attempt to teach find no common ground.

Any educational situation involves a purpose or objective, a content to be learned (a curriculum), and a process of learning. The over-all purpose or objective, what we are trying to do, is expressed as theory of education. The *process* of changing behavior in accordance with educational theory is expressed as the psychology of learning. Between these two is the content to be learned, whether substantive subject-matter or behavior traits. The thesis of this paper is that if the psychology of learning is to be effective in the schools, it must focus its interests and design its experiments with awareness of the theory of education that is currently accepted by the society in which the schools operate. As theory of education is modified, the direction of psychological study must be changed if its results are to influence educational practice. This is no different in the areas of education and psychology than in other areas. The current interest in nuclear physics has for a decade produced a new focus of attention for scientists of many fields. The movement from old issues to new characterizes a live science. Research in the field of educational psychology should reflect the movement of educational theory. In the 1920's, research on drill in arithmetic was in harmony with educational theory; in 1955, such studies would have far less significance because the purpose and theory of teaching arithmetic have undergone marked change.

The ultimate objective, of course, is not merely to relate research to currently accepted educational theory, but rather, to do the kind of research that will help to formulate an educational theory that is *valid*.

I should like to illustrate this relationship between educational theory and the psychology of learning by reference to only two of the several theories of education which might be described. During the last century, and in some places still, education was considered as basically the acquisition of knowledge. The inscription, "Knowledge is Power," was found on many school buildings. Passing on the intellectual heritage of the race was considered the primary purpose of the school. Acquirement of knowledge, with some attendant skills, was the basic aim. Prior to the Civil War, the accumulation of knowledge was not so great as to constitute any considerable burden to the learner. Then the situation changed. Particularly in the field of science, a great expansion began. This increase in subject-matter to be learned was met by devising the elective system. This gave rise to the problem of what knowledge is of most worth and resulted in a fight for time among the different branches of the curriculum. By the twentieth century the situation began to be serious, particularly in the secondary school and junior college. Pupil programs, which at first were made up of a large common group of required subjects plus a few electives, began to show a progressively smaller group of required subjects and an increase in the electives. By the 1920's it became clear that any attempt to cover the field of essential knowledge was hopeless by the required and elective system. A new venture into general education, by means of survey courses, was tried, and now such courses are a part of the program at the junior college level in many institutions. Back of all these changes was the basic question, "What is an education?" As long as it is defined as mastery of the body of knowledge, it must necessarily expand, although the individual learner's capacity to learn has undergone no corresponding change. The elective program was characterized as a specialization in depth in a few areas, with chasms of ignorance between. After thirty years of experimenting with survey courses, they are often subject to the charge of specialization in another axis; namely, specialization in superficialities. But if in theory a good education is defined in terms of knowledge and skill, what other alternatives

are there, particularly in the face of a certain increase in the rate of accumulation of knowledge.

If some modification of the foregoing picture is accepted as a theory of education, then what type of psychological research in learning will likely follow? With so much to be learned, one would expect emphasis on the psychology of memory and retention. Problems of length and distribution of practice periods, whole and part methods of memorizing, prerequisites affecting the order in which different subject-matters should be learned would seem important. Major concern would be centered on subject-matter to be learned, rather than changes in the learner. One can find many specifics to illustrate both this theory of education and the kinds of learning studies mentioned.

However, one might conceive of a second very different theory of education. Rather than define a good education in terms of the knowledge and skills possessed, one might think of it in terms of certain changes brought about in the learner, by means of which he would develop abilities enabling him to respond successfully to the diverse and unpredictable situations that life will bring. Advocates of such a view would readily admit the necessity of a fairly large amount of basic knowledge and skill. Certainly they would not make a virtue of ignorance. But rather than conceive of education as an encyclopedic coverage of knowledge, they would deal with the heritage of knowledge in a highly selective way. They would probably agree on the necessity of certain knowledge and skill, for example, ability to read, knowing how to spell, knowing number facts and relations, and the skill of writing, plus other less well agreed upon additions. However, they would stress such general outcomes as learning how to think, learning how to use a library; how to design and carry out an experiment; how to meet people; how to control the emotions; how to adjust to frustrations. In the literature, the outcomes just enumerated are often dealt with in a very nebulous manner. However, if a person holding some such theory of education were asked to epitomize it, he might answer somewhat as follows: The expanding body of knowledge has already reached such proportions that an attempt by an individual to cover it is hopeless. Therefore, a better procedure might be to select carefully and coherently from racial experience a basic body of knowledge and skills which would serve as tools and back-

ground for whatever kinds of learning one might need to acquire. Using this carefully selected body of facts, concepts, skills, experiences, which would probably constitute the main load of early education, the learner would then test and enlarge his abilities through a rather thorough learning of some sample fields. The essence of such a theory of education is that it is the function of the school to help the pupil learn how to learn; that, beyond the acquiring of basic knowledge and concepts, the purpose of the school is to provide some excellent samples of learning experience in a number of fields and extending over the usual number of years devoted to general education. On completion of such a period of education, the graduate might say, "There are many things I do not know and many skills I do not have, but I know how to get them, I know how to learn what needs to be learned." This concept is the opposite from blueprinting and stereotyping the education of a child. It aims at giving him versatility and independence.

Now, supposing one held some such theory of education as has just been described, in what kind of a psychology of learning would he be most interested? Obviously, the crucial psychological problem for such a theory is that of transfer. How can the learning attained carry over and spread to other situations as they are encountered? How can the outcomes of education be generalized so as to be broad in scope? Certainly the early types of experiments on transfer, such as Thorndike's study of the effect of practice in judging the size of circles upon ability to judge the size of squares, would contribute little. The problem could not be met by studying the carry-over of one academic subject to another subject. Rather, the studies that have significance are those that deal with transfer at the general rather than the specific level; with the development of intellectual habits that may spread widely, rather than with narrow intellectual functions. Studies such as Harlow's experiments with chimpanzees in learning how to learn have high significance for such a theory of education, and they would have still higher significance if carried on with human subjects where the possession of language enhances the possibilities of transferring training.

The concept of transfer in learning how to learn applies both to subject-matter and to method and technique. The understanding embodied in the arithmetical generalization that "when both the numerator and denominator of a fraction are multiplied by the same number, the value of the fraction is not changed" is more

transferable than the specific fact that two-thirds and four-sixths have equivalent value. Likewise, the acquisition of general techniques, such as how to use a library card catalogue, is of more general value than to learn the location on the shelves of a specific reference. The essential test of any theory of education is its transfer value, yet the design of most of our learning experiments dealing with transfer is feeble compared to the size of the problem to be studied.

The crux of my contention is that educational psychologists have been loath to strike out independently in solving problems in their own field. In the main, our problems deal with education in schools, with human subjects, and with higher mental processes involving language. Without the slightest criticism of experiments in general psychology, we cannot continue to be satisfied with implications for education from results of experiments with simple mental processes, with animals, and at the sub-language level. This past summer, parents of children from coast to coast were aroused by a recent book denouncing current methods of learning how to read and proposing that the schools go back to an earlier phonetic method as the sole way of learning to read. The answer to the place of phonics in learning to read will never come from animal experiments nor from the use of nonsense syllables. Nor will it come from half of the educational experiments on phonics which measure only the effect on the ability to pronounce words. If by reading one means the ability to get meaning from a printed page, there is little experimental evidence one way or the other at the present time regarding the value of phonics. The field of reading has without doubt been subjected to more careful research than any other area of learning in the school, yet here in the case of phonics is a prime example of an opportunity passed by to do research on an important transferable function.

I am proposing that educational psychologists take their cues for research from problems of learning in schools, where the processes are complex and where the learnings carried on are at the language level. May I try to indicate what, to me, seem to be three very fruitful fields for research.

I would suggest, first, research on the success, or lack of it, in teaching students how to think. For some years our schools, particularly at the high school and college level, have proclaimed this as one of their main objectives. Yet, a critical appraisal of avail-

able research on this problem gives little evidence that schools are accomplishing their objective. When allowance is made for constantly increasing mental maturity, and when pertinent variables are controlled, it is difficult to find evidence of any marked gains in ability to think due to the work of the schools. Our claims far outrun the evidence. Is it possible to learn how to think more effectively, or are the methods of the school wrong? There is plenty of evidence that students can learn to think within the area of the subject-matter being taught, but the real issue is whether or not they can transfer the ability to other areas. For example, in a study of problem-solving last year, we found that in one group of sixty-one college students, all of whom had had a course in algebra, only twenty of the sixty-one attempted to use algebra in solving a problem where the algebraic method was the most economical way to solution, and, of the twenty who attempted to use algebra, only one succeeded in getting the correct answer. Algebra had apparently failed to contribute a method of thinking to these students. Or, to illustrate further, it is probably true that most students, and some teachers, fail to sense that the high school course in geometry is in essence a course in deductive logical thinking, using space figures as the medium of operation. Rather, geometry is usually taught as a body of content, the learning of which is a sufficient outcome in itself. Here is a major problem in learning at the abstract language level, yet the bulk of the experimentation on it is with animals and at the sub-language level. Why do we so generally avoid experimenting on problem-solving with human subjects and in school situations where the real issues are?

I would suggest that a second area of research intimately related to current educational theory is in the learning of personality characteristics. The school expects its pupils to learn to be accurate, to carry responsibility, to be critical, to have good personal relations with others, and many more such traits. Here, clearly, is a problem in transfer of learning. Most schools deal with this problem of personal-social education in a loose, incidental manner. Whereas the school has a definite and coherent methodology, supported by a considerable body of research, for learning to read and to do arithmetic, there is little experimental evidence to guide the school in its program of personal-social education. In a widely used book on the subject, there is a bibliography of more than eleven hundred references dealing with personality; but less than

a dozen of them report studies of learning personal-social traits in the schools. Admittedly, experimentation in schools is difficult, but can we not design studies with sufficient controls to take account of the difficulties?

Last, I would suggest the problem of motivation in school as of high importance for educational psychologists. Here, clearly, we are in difficulty if we rely on implications from studies of animal learning. In most animal studies, the motivation for behavior is hunger or punishment. Such incentives, obviously, are not applicable in the schools. In fact, the whole concept of tissue needs has little force for school learning. Most of the interests and motives that operate in the schools are learned, and it may well be that one of the main contributions of the school is to modify, through learning, the motivation that operates within the learner. If, for example, the school could succeed in making students intellectually curious, the operation of such a motive might have a broader effect than the mastery of any specific segment of subject-matter. The problem is, how and why do some students learn to be intellectually curious while others do not.

The three problems for research just suggested, namely, learning how to think, learning personal-social traits, and learning motivations, are not matters of great concern to a "Knowledge is Power" type of educational theory. However, to a theory of education based on the concept that the essence of education is learning how to learn, such problems are at the heart of the matter. With such a theory, the main concern is to find what curricular content and what intellectual tools are essential to carry on the kind of learning desired, and then, in turn, to what scope of subject-matter and to what depth and thoroughness should the school go in order to be sure that the student will be independent and versatile, and will feel confident that he knows how to learn as new occasions may require.

Educational psychologists must first of all be psychologists; there is no substitute for competence in the basic subject. But their field of operation is education and, in the main, is concerned with the schools which are society's principal agent for providing education. The main business of the school is learning, and psychologists should have something to say about learning. Whether what they say is understood and is useful to the school depends a great deal on whether or not the psychologist understands theories of educa-

tion as well as theories of learning, and whether his research deals with problems which are important to the educational theory of the time or whether it deals with yesterday's issues. This dual rôle of the educational psychologist is not easy, but it is necessary.

The proposal of this paper in no sense suggests that educational psychologists do only applied research and leave basic research to the general psychologist. But basic research need not be restricted to simple mental processes, nor to subjects at the sub-language level. Basic study of motivation can deal with learned motives as well as physiological needs. Studies of transfer can deal with situations that operate through generalizations that are expressed in language symbols as well as with mazes and puzzle boxes. The designs of research will have to be more complex, but they can be made so. If educational psychologists can play their rôle at this higher level of research, they will be able to contribute significantly to the validation of educational theory.

REFERENCES

- 1) E. R. Guthrie, "The Psychology of Learning," in *National Society for the Study of Education*, Forty-first Yearbook, Part II, 1952.
- 2) E. R. Guthrie, *The Psychology of Learning*, Revised Edition, pp. 310. New York, Harper & Brothers, 1952.

COLLEGE TEACHING ABILITY, SCHOLARLY ACTIVITY AND PERSONALITY

A. H. MASLOW and W. ZIMMERMAN

Brandeis University and San Diego State College

The studies here described were begun out of skepticism with the common tendency to dichotomize teaching and research (or creating), as if they were different, antagonistic, and mutually exclusive, and as if a professor had to be a teacher *or* a researcher. The question we tried to answer was, "Is creativeness (research, activity in the field, writing) in a college teacher positively or negatively correlated with goodness of teaching?" This amounts to asking, "Is the good college teacher more apt or less apt to be active also in his chosen field?"

This research was carried through by the senior author at a large municipal college with very high intellectual and academic standards during the years 1943-1946. Both students and faculty members were used, the former for rating their instructors, the latter for rating colleagues in their own department. Students were asked to list all the courses they had taken up to the date of rating, as well as the instructors teaching those courses. They were then instructed to rate each instructor first "as a teacher" and then "as a personality," defining these terms in their own way. Ratings ranged from +2 for "very good" to -2 for "very poor." The form on which this information was recorded also asked for the student rater's sex, age, class, major subject, and minor subject, but not his name. The raters were assured that these inquiries were being made for research purposes only, and that the data would never be seen by anyone but the experimenter until published in impersonal form.

Faculty members were given a list of their department colleagues, along with a detailed and carefully prepared instruction sheet. Approximately an hour was spent in individual discussion with each rater, explaining the purposes of the research, allaying suspicion, making guarantees, repeating instructions, etc. The faculty raters were asked to rate each of their colleagues as "teachers," as "personalities," and for "creativeness in the field."

The complete instructions, can be briefly characterized as fol-

lows:

A person deserving the highest possible rating, as a teacher, was described as one who is both capable and efficient, who loves his job and manages to inspire his students, who is himself inspired with his work, who is talented, and who not only respects and appreciates his students, but also has good relations with them. The highest rating as a personality was to be given to a very healthy, well integrated person, subjectively at ease with himself, happy or content, using all his constructive capacities, enjoying life without neurotic or psychotic maladjustment. Creativeness was described as aliveness and activity in the chosen field of knowledge, having ideas, continuing to learn and to study, producing something in terms of publications, artistic contributions, or inventions.

Rating data were included for analysis if a teacher was rated by at least three colleagues (median number 3.5), and if the number of teachers so rated constituted a majority in the department. Not less than eleven students rated each of these teachers (median number 35). Complete data were collected on eighty-six teachers who comprised better than sixty percent of the faculty in the following departments of the College: Biology, Chemistry, Mathematics, Psychology, Political Science, and Speech.

Estimated reliabilities of the ratings are listed in Table I, while intercorrelations among the five ratings are listed in Table II.

The pattern of the intercorrelations shows the two student ratings, "as a teacher" and "as a personality" holding together, and the two colleague ratings, "as a teacher" and "for creativeness" doing likewise. At the same time, students and colleagues agree unexpectedly well in their ratings "as a teacher," but the colleague rating "as a personality" stands apart from all other ratings by either colleagues or students. This clustering becomes clearly apparent in a rough factor-analytic approximation.¹

¹ Centroid method with rotation to simple structure and positive manifold, yielding the following factor loadings:

Teacher's rating of colleagues "as a teacher."	0.71	0.44	0.40	0.85
Teacher's rating of colleagues "as a personality."	0.34	0.13	0.66	0.57
Teacher's rating of colleagues "for creativeness."	0.77	0.25	0.35	0.77
Student's rating of teachers "as a teacher."	0.10	0.77	0.45	0.80
Student's rating of teachers "as a personality."	0.30	0.82	0.12	0.77

TABLE I—RELIABILITIES* OF THE RATINGS
($N = 86$)

Rating	Estimated Reliability	Estimated Reliability Corrected	Standard Error of the Estimated Reliability Coefficient
Teacher's ratings of colleagues "as a teacher"	0.61	0.85	0.040
Teacher's ratings of colleagues "as a personality"	0.47	0.77	0.049
Teacher's ratings of colleagues "for creativeness"	0.63	0.86	0.038
Student's ratings of teachers "as a teacher"	0.77	0.96	0.044
Student's ratings of teachers "as a personality"	0.70	0.94	0.055

* Ratings of a teacher submitted by three colleagues chosen at random were plotted in pairs and correlated, while the averages of five student raters chosen at random were correlated with the averages of a second five chosen at random. Coefficients so obtained were corrected by entering the ratio of the number of ratings averaged to the median number of raters in the Spearman-Brown prophecy formula.

Thus, on the basis of the data presented, it can be concluded that, at least in the college where the ratings were obtained, colleagues tend to equate good teaching with creativeness, and students tend to equate good teaching with good personality. Also, students and faculty agree fairly well on who the good teachers are, but their conception of personality is quite different. In fact, the faculty conception of personality is relatively independent of all of the other rated traits.

This study is open to criticism on several counts. In the first place, ratings are subjective evaluations at best. Moreover, two populations were used as raters, one of which was instructed in detail while the other was given minimal instructions.² Also, the detailed instructions given to one group of raters were so worded as to introduce automatically a certain amount of overlap. These facts are regrettable; yet, some of the findings may be seen as even more remarkable in this light.

² The two sets of rating data were originally gathered separately. The decision to combine them was made later when it appeared that, despite inconsistencies, useful conclusions might be drawn from the comparisons.

TABLE II—INTERCORRELATIONS* OF THE RATINGS
($N = 86$)

Variables Correlated	r	Standard Error of r
Colleague's ratings "as a teacher" vs. colleague's ratings "as a personality"	0.58	0.072
Colleague's ratings "as a teacher" vs. colleague's ratings "for creativeness"	0.77	0.044
Colleague's ratings "as a teacher" vs. student's ratings "as a teacher"	0.69	0.057
Colleague's rating "as a teacher" vs. student's rating "as a personality"	0.59	0.071
Colleague's rating "as a personality" vs. colleague's rating "for creativeness"	0.51	0.080
Colleague's rating "as a personality" vs. student's rating "as a teacher"	0.29	0.099
Colleague's rating "as a personality" vs. student's rating "as a personality"	0.40	0.091
Colleague's rating "for creativeness" vs. student's rating "as a teacher"	0.51	0.080
Colleague's rating "for creativeness" vs. student's rating "as a personality"	0.41	0.091
Student's rating "as a teacher" vs. student's rating "as a personality"	0.76	0.046

* Pearson product moment coefficients were computed. Distributions were markedly skewed with too few raw score categories for normalizing, but the correlation plots showed linear regressions. The skewness was attributed to faulty measuring scales, rather than to non-normal population distributions.

It would seem that if the high relationship between two of the colleague ratings was due primarily to the operation of a halo effect in rating, or to overlap introduced by the instructions, the third colleague rating would be equally affected. Yet it stood well apart from the others. Also, if the two rater populations received different rating instructions, the intercorrelations between ratings by the two populations would be expected to be lowered. But the correlation between their ratings "as a teacher" was quite substantial. These facts seem to strengthen rather than weaken the conclusions drawn above.

It is not claimed that these results would necessarily obtain in other institutions. Brooklyn College is unusual in two respects that help to create these results. First, its students are homogeneously

excellent and serious. They take their teachers and courses very seriously, and are apt to feel swindled rather than relieved by an empty course. Though, like other students, they may choose instructors who are popular, entertaining, or well liked, they do not necessarily respect them as teachers. Second, the faculty of this college is very heterogeneous, ranging from excellent to incompetent. The senior author's personal judgment is that the best twenty-five per cent would grace any faculty in the country, but that the poorest twenty-five per cent have no real usefulness to the college, since their ability and intelligence are probably less than that of most of their students.

A faculty which is homogeneously good (or homogeneously bad) undoubtedly would show lower correlations between teaching ability and creativeness. Our findings would be much less striking in such a faculty. In addition, we should expect student ratings of their teachers to be useful only when the students were earnest and serious intellectually.

Finally we may say a word about the validity of student judgments of their teachers. These correlate so well ($r = 0.69$) with faculty judgments of these same teachers that a faculty cannot take student judgment lightly without casting aspersions on its own competence to judge.

BOOK REVIEWS

HELEN D. SARGENT. *The Insight Test* (a Verbal Projective Test for Personality Study). The Menninger Clinical Monograph Series, No. 10. New York: Grune and Stratton, 1953. \$6.75

Helen Sargent's book is essentially a technical manual for use with a new projective device, the Insight Test. The Insight Test should be of interest to clinical psychologists who wish to add to their armament one more weapon for attacking the ego structure of their clients. However, the book and test will be of little value to psychologists or educators who are not already thoroughly familiar with dynamic psychology and the theory and practice of projective techniques.

Sargent is modest in her claims for the Insight Test and suggests that it be used as a supplement to, rather than a substitute for, the widely used projective tests such as the Rorschach method and the Thematic Apperception Test. Furthermore, she insists that her test is still in an experimental stage and can, thus, be used only with caution by persons well trained in clinical psychology.

Although Sargent's Insight Test has a number of characteristics in common with well-known projective tests, it differs from them in its stimulus material, scoring system, and interpretation.

The stimulus material is a series of verbal items called "armatures," a term derived from the flexible frames or skeletons on which an artist constructs his models. These armatures give the bare outlines of a problem situation, and the subject responds by telling what the leading character did, and why, and how he felt about it. The rationale for labeling this task "The Insight Test" is that the subject is, ostensibly, asked to show insight into the behavior of others. Actually, of course, it is the clinician who hopes to gain "insight" into the ego structure of the subject.

A large part of the book is devoted to the fairly complex scoring system and to illustrative protocols. The author claims that, although the scoring system appears complex and time-consuming at first, it can be easily learned and need not take more than thirty or forty-five minutes per record, as a rule. Even so, this represents an expenditure of time and effort that the busy clinician may be reluctant to spend on a new and still frankly experimental test, especially when he is told that the Insight Test is best used as a

supplement to such time-consumers as the Rorschach and T.A.T.

Interpretation of the test, once scored, may follow quantitative or qualitative data derived from the protocols and will vary to some extent with the particular clinician using the test. However, Sargent's orientation is psychoanalytical, in a broad sense, with emphasis on ego psychology and dynamic theories of personality. A formal analysis is used with the Insight Test, and the interpretation is concerned with effect and thought processes that can be observed in the solution of "armature" problems.

The problems of reliability and validity, in any traditional sense, are always hard to cope with in a projective test for the study of personality. Sargent reports some predictive studies are in progress, but there is little "validity" data. She reports fair agreement on interpretation between users of the test and talks at some length about "intrinsic validity".

Reception of the Insight Test will depend upon the sympathies or antipathies of the psychologist reviewing or employing the test. To a clinician who is sympathetic to projective testing and who is willing to do a good deal of work and experimentation himself, Sargent's book will prove to be worthwhile reading. To the psychologist who is concerned largely with normal or non-pathological behavior or who has reservations regarding projective psychology, the book will be read, if at all, with skepticism. What is needed now is more research evaluating the predictive usefulness of the Insight Test in diagnosis and therapy.

HENRY H. MORGAN

The Psychological Corporation

EDWARD PODOLSKY, Editor. *Music Therapy*. Philosophical Library. New York: 1954, pp. 335.

Music Therapy, edited by Edward Podolsky, contains thirty-two articles by thirty-nine contributors on almost every phase of music as related to mental health and therapy that can be considered. The articles could be grouped into three large sections: a series of articles on the general value of musical therapy and the history of it; the effect of music on moods and temperament; and, another series, a larger series, on musical therapy in specific assignments—for example, the use of music on psychotic patients, the use of it in anxiety as an adjunct on electroshock therapy, the use of it in anxiety states and depression, tension headaches, psychopathic personality,

gastric disorders, high blood pressure, the heart, and, to be up to date, it has a chapter on music in military medicine. Also, as a sign of modernity, it has a small article on high fidelity in music therapy. In general, the articles are written on a descriptive or vague generalized level. Relatively little objective data are given, although some of the programs used in some of the institutions are described and some of the results in general terms given. A fair number of these articles were previously published in about fourteen periodicals listed and acknowledged. There is very little editing done.

The articles do not hang together; they are not arranged in any orderly fashion; they are just a series of relatively small articles—typically, by one person, every once in a while by three or four. There are some objective data in a few of the articles; for example, in *The Effects of Personal Insecurity on Reactions to Unfamiliar Music*, some simple data are presented and generalized from. The last article by Pepinsky does at least suggest limitations of research in music therapy, but even this is on the vague level. It starts off by again referring to the fact that the Bible makes more than casual mention of the use of music as therapy, ends with the advocacy of the use of Lloyd Morgan's Canon, and refers to the dilemma which confronts the researcher in music therapy who finds himself either doing clinical observing without controlled experimentation or attempting basic research without consideration of the interaction of the total personality and the environment of the patient.

Musical therapy is a reality. The description of the use of it in a number of the articles included in this book would be of interest and value to people who are in position to use it, but, as far as objective research is concerned, based on any kind of experimental design that is realistic and discriminating, the beginnings are still to be made. But the volume as is—sketchy articles by a large number of contributors about a large variety of subjects with plenty of overlapping and generalized references—does contain material of value for people in a position to use music as therapy in institutional work or in other situations which permit it. The book contains a bibliography but no index.

H. MELTZER

Psychological Service Center, St. Louis, Mo.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

April, 1956

Number 4

A STUDY OF STUDENT TEACHERS' ATTITUDES TOWARD CHILDREN¹

MERRITT C. OELKE

School of Education, The Florida State University

This study was a preliminary exploration of the nature of some of the changes in student teachers' attitudes toward children during initial teaching experiences and was designed to locate promising areas for further study.

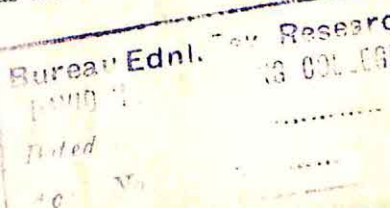
The Minnesota Teacher Attitude Inventory (1) and a modification of the Alexander Thematic Apperception Test for Teachers (2) were used as stimuli to elicit expressions of attitudes. The former is a conventional inventory designed to measure teachers' attitudes toward children, while the latter consists of eight pictures which may be interpreted by the subjects as containing situations involving children.

The modifications of the Thematic Apperception Test for Teachers in this study consisted of the usual T.A.T. instructions for five pictures used previously by Ohlsen and Schultz (3), plus the use of a series of questions based on kinds of responses which those investigators found discriminated between best and poorest student teachers. The questions were designed to keep structuring minimized.

A total quantitative score using expressed acceptances and rejections of children's behavior was developed, so that data could be analyzed statistically by means of simple analysis of variance, chi-square, and the "t" test.

No extensions of this research to other groups were intended since the responses were both the criterion and the test. No generalizations of the behaviors of other groups nor extension of these re-

¹ This report is based on a part of a Doctor of Education thesis at the University of Illinois. Dr. Merle M. Ohlsen was the adviser.



sponses to infer performance with children should be made without further specific validation.

The professional semester at a large midwestern university furnished the necessary and desirable experimental conditions for measuring the changes studied. In this program the student teachers were on campus the first six and the last four weeks of the semester, while the middle six-week period was spent away from campus teaching in a community.

A group of forty-four senior student teachers and a group of fifty junior students in education were used in a test-retest procedure. The seniors were members of two educational psychology classes in the professional semester. Since most common teaching fields were represented and since the only basis for enrolling in a particular section was that of scheduling conflicts, random assignment was assumed. The juniors were students in a course in principles of secondary education and did not differ from the seniors in areas essential to this study.

The two tests were administered to the teaching group on successive days during the week prior to the beginning of student teaching and again seven weeks later during the first week following student teaching. The T.A.T was administered to the junior pre-teaching group twice with exactly seven weeks between administrations. The M.T.A.I was not used for this group.

Correlations between T.A.T testings were 0.62 for the pre-teaching group and 0.55 for the teaching group. While no statistically significant differences were found between the two groups on the testings as far as total score was concerned, significant differences in amount of change were located in some areas.

A change significant at the one per cent level was found for the pre-teaching group in the direction of greater expressed acceptance of dependency upon adults and nonachievement as normal behaviors of children. This may be a result of the kind of course work these subjects were doing since the course in principles of secondary education at this university tended to emphasize the rôles of the teacher and student and to develop a greater understanding of the difficulties of a child. On the other hand, the teaching group showed less tendency to select the child as the most liked person after student teaching. Significance was at the five per cent level. This apparent contradiction may be accounted for by considering acceptance and liking as somewhat different functions.

TABLE I.—SUMMARY OF CHI-SQUARE TESTS OF SIGNIFICANCE OF CHANGE IN SCORING AREAS

Scoring Area	Pre-teaching Group	Teaching Group
Pleasure	2.5491	0.0043
Sexuality	0.4726	0.0054
Aggression	0.0644	0.3797
Dependency	6.8727*	0.0254
Nonachievement	12.4500*	0.5640
Affection	0.5963	0.0296
Curiosity	a†	a†
Interpersonal	1.7545	1.5359
Empathy	0.7364	4.5630†
Democratic	0.0180	0.0316
Optimistic	0.2104	12.6577*

* Significant at 0.01 level, $df = 1$.

† Expected frequencies less than five.

‡ Significant at 0.05 level, $df = 1$.

The teaching group produced more optimistic outcomes in their protocols after student teaching than before, with the change significant at the one per cent level.

Neither of the instruments used in this study demonstrated any statistically significant differences which could have been attributed to teaching field, age, sex, school where student teaching was done, previous informal teaching experience, degree of control by coöperating teacher, student teachers' perception of difficulty, number of teachers in the family, number of siblings, or membership in a particular section of educational psychology. In about half of the analyses, the within groups variance was greater than the between groups variance so that F was not computed. In the balance, the F had a probability far greater than the five per cent level.

There was a suggestion in the data that in home economics and social studies, student teachers may be more acceptant of children than in other fields although this is only indicated. This should be investigated further since these two groups were the only ones in the study which had a planned program of pre-student teaching experiences. Another suggestion which may warrant further investigation is based on the student teachers' perception of difficulty. Those who felt they had no difficulty in teaching tended to increase slightly in acceptance; those who felt a little difficulty had no score change; and those who felt they had great difficulty tended to de-

crease in acceptance of children. Neither of these suggestions is based on statistical data but both seem to warrant further investigation.

Some question might be raised concerning the representativeness of the norm group for the Minnesota Teacher Attitude Inventory since the original norms were based on University of Minnesota students and the academic group in the present study obtained scores significantly different (five per cent level) from the mean of the norms. Test-retest product moment correlation was 0.88 for the present group. This coefficient was as high or higher than correlations reported by the authors for various time intervals.

There seems to be little relationship between the T.A.T and M.T.A.I. measured, since the product moment correlations were -0.17 on the pre-test, $+0.24$ on the post-test, and $+0.15$ on score change.

Since mean scores may tend to conceal significant information, a study was made of the scores of the fifteen per cent best student teachers and the fifteen per cent poorest student teachers as selected by their supervisors (4). The numbers were necessarily very small and results should be considered only as indications which point the way for further study.

The Minnesota Teacher Attitude Inventory did not show any statistically significant differences between the two groups.

The T.A.T seemed to be of value primarily in terms of change recorded for these groups since amount and direction of score change differed to a degree which was significant at the five per cent level. Table III indicates the difference in direction of change with the

TABLE II.—COMPARISON OF MEAN MTAI SCORES OF STUDY GROUP AND STANDARDIZATION GROUP

Group	N	Mean	S.D.
Academic subjects:			24.3
Standardization	237	67.8	21.0*
Study pre-test	22	55.4*	29.5*
Study post-test	22	54.8*	
Nonacademic subjects:			25.4
Standardization	238	63.3	18.8
Study pre-test	22	60.3	28.9
Study post-test	22	62.4	

* $P < 0.05$.

TABLE III.—RAW SCORE CHANGE SUMMARY ON TAT AND MTAI
BEST AND POOREST STUDENT TEACHERS

Test	Group	N	Mean	Range	F	Probability
TAT	Best	6	+9.3	-2 to +29	6.76	$P < 0.05$
	Poorest	7	-6.9	-25 to +5		
MTAI	Best	6	+1.5	-11 to +16	F not computed*	
	Poorest	7	-0.7	-15 to +13		

* Within groups variance greater than between groups variance.

positive score demonstrating an increased acceptance as measured on this instrument. It is perhaps significant that the only increase in score in the poorest group was registered by a subject who was placed there because of difficulties in relations with the coöperating teacher rather than because of difficulties with students.

While it was impractical to compute statistical significance, a pattern analysis of the T.A.T scoring indicated that the best group seemed to be more accepting of pleasure, aggression, and affection as normal behaviors of children and seemed to express a greater liking for children and a greater desire for interpersonal relations with children. Further investigation seems warranted by these differences.

The following conclusions seem sufficiently clear to warrant further study:

(1) Significant differences were found between teaching and pre-teaching groups in expressed acceptance of aggression, in selection of a child as most liked, and in optimism. The teaching group was more accepting and more optimistic.

(2) The T.A.T discriminated in amount and direction of change between the small number of best and poorest student teachers while the Minnesota Teacher Attitude Inventory did not.

(3) Student teachers rated as best expressed greater acceptance of children's behaviors after student teaching. Those rated as poorest tended to become less acceptant of children's behaviors after teaching.

(4) The Minnesota Teacher Attitude Inventory and the Thematic Apperception Test for Teachers as modified for this study measured different characteristics of the student teachers.

(5) Neither instrument used showed differences on the pre-test sufficiently significant to warrant use of the tests as predictive in-

struments at this stage of development. This is particularly true when used on a one-administration basis.

(6) The norm groups of the Minnesota Teacher Attitude Inventory as used in this study should be examined for representativeness.

(7) Planned pre-student teaching experiences should be studied further to determine if there is a definite relationship between these experiences and the kind of relationships established with students while teaching.

Finally, these and so many other questions were raised by this study that there is ample reason to believe that a thorough study of student teacher attitudes will be most productive for future research.

REFERENCES

- 1) Walter W. Cook, Carroll H. Leeds, and Robert Callis, *Minnesota Teacher Attitude Inventory*. New York, Psychological Corporation, 1951.
- 2) Theron Alexander, *The Prediction of Teacher-Pupil Interactions with a Projective Test*, Ph.D. Thesis, University of Chicago, 1949, 66 pp.
- 3) Merle M. Ohlsen and Raymond E. Schultz, "Projective Test Response Patterns for Best and Poorest Student Teachers," *Educational and Psychological Measurement*, Vol. 15, No. 1: 18-27.
- 4) Merle M. Ohlsen and Raymond E. Schultz, "As Best and Poorest Student-Teachers Are Seen by Their Peers," *Educational Administration and Supervision*, Vol. 40, No. 1: 22-23.

A STUDY OF STUDENT WITHDRAWAL AT "W. U."¹

EMMA W. BRAGG

Professor of Education, Texas Southern University

STATEMENT OF PROBLEM

The problem of the study consists of testing the null hypothesis that there are no statistically significant differences in the performance of those students who withdrew and those who continued on certain selected criteria considered related to withdrawal. The criteria chosen against which the two groups of students are compared are: High School Average; Age; Performance on the American Council on Education Psychological Examination (*Q*, *L* and *T* Scores, 1948 College Edition); Coöperative English Test, *C*₂, Comprehension, Higher Level; Academic Average made at "W. U." during the first semester of 1952-53; Sex, and Curriculum.

SELECTION OF SAMPLE

The defined population is the freshman day class that entered in the fall of 1952. This class contained seven hundred and twenty-four students; three hundred and five or 42.13 per cent continued and four hundred and nineteen or 57.87 per cent withdrew during the two-year period from September 1952 through May 1954. Random samples of 100 cases were taken, with the use of the table of random numbers from each of the two groups of students.

DESIGN AND STATISTICS

The statistical design is that of the comparison of unmatched groups. The statistical technique of Critical Ratio was used to test the differences in mean performance between the "continued" and "withdrawn" groups except for the entrance test data which were tested for differences in per cent scoring above the selected percentiles of twenty, fifty, and seventy-five (national norms used). Chi-square was used to determine if there were any statistically sig-

¹ "W. U." is a midwestern city college with a student population of two thousand, nine hundred and nineteen; coeducational; inter-racial; schools of arts and sciences; education; fine arts; business administration; graduate.

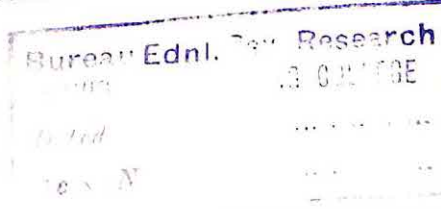


TABLE I.—COMPARISON OF THE TWO GROUPS ON SELECTED CRITERIA OF WITHDRAWAL

Criterion	"Con- tinued" Group	"With- drawn" Group	Diff.	S.E. Diff.	C.R.	P
1. High school aver- age	87.59	85.05	2.54	0.591	4.30	<0.003
2. Mean age	18.73	19.12	0.39	0.361	1.08	0.29
3. Entrance test scores (per cent scoring above)						
A. A.C.E.						
1) "I"						
20%ile*	80.00	62.00	18	6.30	2.86	0.004
2) "I"						
50%ile	42.00	34.00	8	6.80	1.18	0.25
3) "I"						
75%ile	20.00	18.00	2	5.54	0.36	0.73
4) "Q"						
20%ile	80.00	75.00	5	5.89	0.85	0.40
5) "Q"						
50%ile	58.00	47.00	11	7.02	1.57	0.12
6) "Q"						
75%ile	35.00	23.00	12	6.36	1.89	0.06
7) "Total"						
20%ile	77.00	63.00	14	6.40	2.19	0.03
8) "Total"						
50%ile	51.00	34.00	17	6.88	2.47	0.01
9) "Total"						
75%ile	26.00	22.00	4	6.03	0.66	0.52
B. Coöperative English test						
1) "Total"						
20%ile	73.00	56.00	17	6.65	2.56	0.01
2) "Total"						
50%ile	39.00	20.00	19	6.31	3.01	0.003
3) "Total"						
75%ile	20.00	13.00	7	5.23	1.34	0.19
4. Mean grade†	1.45	0.62	0.83	0.12	6.80	<0.003

* %ile = percentile.

† C = 1; D = 0.

TABLE II.—RELATIONSHIP BETWEEN CONTINUATION:
SEX AND CURRICULUM

Criterion	Chi-square	Degrees of Freedom	P
5. Sex	1.13	1	0.30
6. Curriculum	13.80	5	>0.02

nificant relationship between sex and continuation and curriculum and continuation.

FINDINGS

See Tables I and II.

SUMMARY OF FINDINGS AND GUIDANCE IMPLICATIONS

- (1) There is a statistically significant difference in the performance of the "continued" and "withdrawn" groups on the criterion of high school average, according to Table I. For counseling, the finding may mean that academic counseling should be provided those whose high school average falls much below the mean performance of the "continued" group. Also, the finding may mean that an attempt should be made to select for admission those applicants whose high school average does not fall much below the average of the "continued" group.
- (2) There is no statistically significant relationship between age and continuation at "W. U." However, those in the "withdrawn group" have a higher mean age and a wider range than the "continued" group (see Table I). For counseling, the implications may be that the older students may need more individual attention than the younger students.
- (3) The entrance test data reveal that at "W. U.," emphasis is placed on the linguistic and reading abilities to the neglect of the quantitative ability. The implications may point toward the need for more emphasis on remedial and developmental reading programs. The following are summarizing statements afforded by the entrance test data (see Table I):
 - (a) The "Q" section of the A.C.E. test does not discriminate between those who continued and those who withdrew.

- (b) The "L" section of the A.C.E. test discriminates between the two groups at the twenty percentile but not at the fifty and seventy-five percentiles. Probably those who compared unfavorably on the "L" (linguistic) section withdrew.
 - (c) The "Q" (quantitative) score tends toward statistical significance at the seventy-five percentile but at the twenty and fifty percentiles the "Q" score does not discriminate. It may be that the "Q" section of the test is too easy at the twenty and fifty percentiles as evidenced by the large percentage scoring above these levels (see Table I) and/or that the quantitative ability is not sampled at "W. U." during the first four semesters.
 - (d) The Coöperative English Test total scores discriminate between those who continued and those who withdrew at the twenty and the fifty percentiles but not at the seventy-five percentile. This finding is the same as for the A.C.E. total score. It may be that those who compared unfavorably withdrew so that at the seventy-five percentile, the two groups left were homogeneous.
- (4) There is a statistically significant difference in the performance of the "continued" and "withdrawn" groups on the criterion of grade point average made during the first semester at "W. U.," with the "continued" group making the higher grade point average (see Table I). Withdrawal may be a question of lack of motivation as a result of poor grades or it may be a question of continuation drawing heavily on the reading and verbal abilities.
 - (5) There is no statistically significant relationship between sex and continuance at "W. U." (see Table II).
 - (6) There is a statistically significant relationship between curriculum at "W. U." (see Table II). The implications may be that more academic counseling should be provided in the selection of the curriculum, based upon a matching of skills, abilities, interests, and aptitudes required by the particular curriculum.

AN OBJECTIVE MEASUREMENT OF A TEACHER'S CLASSROOM INTERACTIONS

JOHN WITHALL

National Education Association

It is generally agreed that, after the home, one of the major instruments for socializing individuals in terms of our cultural values and mores is the school. It has also been pretty well determined that this induction into the customs of our society will be carried on in a group situation of twenty-five or more pupils guided by an adult, considered to be a representative of our values, who will oversee and aid the youngsters' learning and growth. We seem committed both from a social, psychological, and economic viewpoint to this method of inculcating in young people the ideas and ideals we have developed over the years. In view of this fact and in view of the "group-guided-by-a-teacher" learning situation which is normal and accepted in our educational system, it seems important to examine the way in which the teacher fulfills his function in the classroom particularly with respect to his distribution of interactions with the individual learners. This we have attempted to do with a regular teacher of an art class.

DISTRIBUTION OF A TEACHER'S TIME AMONG PUPILS

No one would be naïve enough to suggest that a teacher's time should be divided *equally* among the pupils in his class. That would be a negation of our modern view of the peculiar needs of each learner for help, guidance, and support from the teacher. However, an assumption is made in this study that a teacher interested in the maximal growth and development of each pupil will accord each youngster as much time as is needed to render help and guidance effectively. This help and guidance will be provided primarily by face-to-face interactions, the number of which will be governed by the known facts about the needs of pupils that have been gleaned from the teacher's own observations and from tests and school records. To the writer this means that if a given teacher consistently affords few or no face-to-face contacts with a particular

learner, one may legitimately raise the question as to whether in the light of the known facts the teacher is adequately fulfilling his rôle vis-a-vis that pupil. In order to ascertain the pattern of distribution of face-to-face contacts between pupils and teacher, a variation of the Olson time-sampling technique was used. An analysis was made of the photographically recorded contacts of a teacher with his pupils in a normal classroom situation by means of an automatic time-lapse camera. The class was a regular art class. The pupils met daily on a customary schedule for twelve weeks with their regular art teacher. The situation was not experimentally manipulated in any way.

The questions to which answers were sought were the following: How is the teacher's time distributed among the learners? To what extent does the amount of time given to individuals differ? Is the distribution of time consonant with the presumed needs of the pupils? What are the implications for teachers?

HOW THE DATA WERE COLLECTED

For twelve weeks the regular class sessions of an eighth-grade art class in the Laboratory School of the University of Chicago were observed and analyzed. The class sessions were held in the laboratory classroom of the Department of Education of the University of Chicago. The class comprised twenty-six eighth-graders. The pupils worked around four large tables in the art class. They were not assigned specific chairs but moved freely about the room. The teacher went from table to table as he saw fit or was called on. On rare occasions the twenty-six individuals were convened as a whole group by the teacher for special announcements or infrequent class discussions. Each class session was recorded on the sound tape of a Recordgraph, and time-lapse photographs were taken automatically at fifteen-second intervals. As a result it was possible to synchronize the sound with the shots taken at fifteen-second intervals and get a fairly complete view of the interactions and groupings in the classroom.

THE TEACHER AND HIS PHILOSOPHY

The teacher of the class was a friendly, outgoing individual who achieved considerable rapport with his classes. He was considered a "good sport" and was accepted both in and out of class by pupils and colleagues as a "good egg." The educational philosophy of the

teacher is such that his major emphasis is on understanding the learner and trying, insofar as possible, to utilize his skills to meet the needs and harness the basic motivations of the learners. He was not the type of teacher to play favorites in a class but, rather, accepted each individual on his merits and geared his efforts to helping each pupil achieve as full a mastery as feasible of the ideas and skills related to the subject matter.

When the researcher asked the teacher to outline some of the criteria that guided his class activities, he stated that he: (1) was concerned with the development of each pupil's personality; (2) sought to maintain and enhance good pupil-teacher relationships as a means of facilitating learning; (3) was anxious to encourage individual creativity and experimentation; (4) sought to help the pupils identify their own goals and choose their own activities; (5) sought to have the pupils develop some ability in self-evaluation.

It would seem reasonable to expect that the teacher of this art class, activated by such values, would be quite sensitive to each pupil's interests and needs, and that his contacts would be distributed among them on the basis of his appreciation of each learner's special needs and requirements. The evidence gleaned from an analysis of both the sound and film record does not substantiate this reasonable expectation. For instance two individuals, N and F, secured or were accorded a major proportion of his attention, but observations made in and out of the classroom showed healthy peer relationships, better than average ability, healthful spontaneity, and considerable self-sufficiency and self-direction. Hence, presumably, they least needed the help and support of the teacher. An important factor in this seems, in retrospect, to have been the very spontaneity and social aggressiveness of the two. The study was an eye-opener to both the researcher and the teacher in revealing this imbalance of teacher-pupil interactions. This paper presents the data obtained from an analysis of the still photographs of the class taken from eleven randomly selected sessions.

METHOD OF ANALYZING THE DATA

By examining the photographic record of this class on several days and making a careful tally of the pupils who are seen in the pictures gathered around the teacher, it is possible to obtain an accurate representation of the distribution of interactions between

the teacher and the twenty-six pupils. Interactions are here interpreted to mean that the pupil is within the area encompassed by an imaginary circle around the teacher having a radius of approximately one yard. The pupil alone or with others appears in the photograph to be either conversing with the teacher or listening to the teacher's comments addressed either to him or to the group in the defined area. Since the pictures were taken at fifteen-second intervals they are assumed to represent with some degree of accuracy the total pattern of interactions between pupils and teacher. If we take each fifteen-second exposure in turn and tally each child's contact with the teacher, we get a measure of the attention bestowed by the teacher on each member of his class.

BASIC DATA

The basic data offered here represent the contacts between each of the twenty-six pupils and the teacher during eleven randomly selected class sessions from November 6 to December 11. Most of the sessions were seventy-five minutes in length; three were thirty-five minutes. The first eight sessions were "uncontrolled" in the sense that the instructor was not aware of the pattern of distribution of his time and effort among the members of the class. The final three sessions were "controlled" in the sense that the teacher deliberately attempted, in the light of the analysis made of his interactions during the first eight sessions, to redistribute his time and attention among the pupils more equitably.

ANALYSIS OF DISTRIBUTION OF INTERACTIONS IN "UNCONTROLLED" SESSIONS

The "uncontrolled" pattern of contacts occurred prior to the teacher's being aware of the way in which he distributed his time in his class. The two pupils, F and N, had more than double (three hundred and forty-eight) and triple (five hundred and eighty-one), respectively, the mean number (one hundred and fifty-one) of contacts within the entire group. The total contacts of these two individuals with the teacher accounted for more than one quarter of all (three thousand, three hundred and thirty-seven) the contacts of the twenty-two youngsters for whom data are presented. On the other hand, the combined total contacts of eight other pupils with the teacher—A, fifty-one; H, thirty-nine; Q, forty-three; R, fifty-five; V, fifty-four; W, sixty-four; X, forty-six;

and Y, eighteen—all together account for a total only slightly larger than F's and only about two-thirds of N's total contacts. It would appear that despite the teacher's sincere and avowed desire to facilitate maximum learning of each member of his class and to enhance sound teacher-pupil relationships, some pupils appeared to be taking a far larger share of his time and attention than others. Q, for example, who was a fringer in the peer group, was rejected by the boys, had more than her share of pubescent problems, was the youngest in the class, had the highest IQ, was ungainly, overweight and not a member of a well-knit girl clique, had a deep need for his time and attention.

On the basis of the number of contacts registered in the eight uncontrolled sessions representing five hundred and fifty-five minutes, it is possible to compute the mean number of contacts per hour for each pupil. These data are presented in Table I.

The range in mean number of contacts per hour for each pupil (Table I) is from 1.9 to 62.8, i.e., 60.9. Thus from these data it is

TABLE I.—NUMBER OF CONTACTS PER HOUR FOR EACH PUPIL
EIGHT UNCONTROLLED* SESSIONS

Pupils	Contacts per Hour	Pupils	Contacts per Hour
A		N	62.8
B	5.5	O	4.7†
C	34.5†	P	29.3
D	17.7	Q	4.6
E	9.6	R	5.9
F	6.7†	S	20.2
G	37.6	T	22.5
H	24.1	U	12.1
I	4.2	V	5.8
J	26.1	W	6.9
K	4.3†	X	4.9
L	17.1	Y	1.9
M	13.3	Z	18.0
	9.4		
Over-all mean: 15.8			

* The teacher was unaware of the pattern of distribution of his contacts in the class.

† Based on four hundred and forty-five-minute total. Two class periods were missed; one of seventy-five minutes, the other of thirty-five minutes.

‡ Based on five hundred and twenty-minute total. One class period of thirty-five minutes was missed.

apparent that in a sixty-minute class period the teacher would give his attention to N once a minute and only once every half hour to Y.

TEACHER'S CONSCIOUS ATTEMPTS TO REDISTRIBUTE CONTACTS AMONG PUPILS

The conscious attempt by the teacher to spread his attention more evenly within the class came about as a result of a discussion which the researcher had with him during one of their regular staff meetings. At this meeting we discussed the problem of some of the isolates¹ in the class. We were impressed with difficulties facing these individuals some of whom, besides being rejected in everyday situations by their peers, seemed also to be "neglected" by the teacher. It was agreed that both the "fringers" and "isolates" and the "neglects" might be helped if the teacher began to evidence a greater interest in them and their work. The researcher and the class teacher agreed that up to this time little rapport seemed to have been established by the teacher with certain pupils, notably: Q, Y, A, O, W, R, C, and L. The teacher was greatly impressed with findings we had made from an analysis of the classroom interaction and was only too anxious to do what he could to alter the situation that existed. He, therefore, undertook, deliberately, to work more closely with those eight whom we had specifically identified as well as with others in the room whom he may previously have tended to overlook. As a result of the teacher's efforts to redistribute his time so that the "neglects" would get a greater share of his attention, we were able to collect data on several "controlled" sessions. These data are presented in Table II.

ANALYSIS OF DISTRIBUTION OF INTERACTIONS IN "CONTROLLED" SESSIONS

The amount of class time involved in the three "controlled" sessions happens to be exactly one third of the elapsed time in the "uncontrolled" sessions. In the one hundred and eighty-five min-

¹ The isolates were discovered from the information gleaned from recorded, individual, open-ended interviews aimed at assessing peer relationships in this group, since the pupils were too test-wise for the usual sociometric procedures on the basis of our intensive observation of these youngsters throughout the school day in a variety of situations—lunch, gymnasium, social studies, and after-school activities.

TABLE II.—NUMBER OF CONTACTS PER HOUR FOR EACH PUPIL CONTROLLED* SESSIONS

Pupils	Contacts per Hour	Pupils	Contacts per Hour
A	23.7	N	46.4
B	25.6	O	6.2
C	28.8	P	45.4
D	53.7	Q	44.4
E	11.2†	R	12.3
F	72.6	S	18.2
G	Shop‡	T	13.9
H	1.6	U	22.4
I	Shop‡	V	13.1
J	7.0	W	4.8
K	12.0	X	29.8
L	35.3	Y	8.1
M	31.5	Z	27.6
		Over-all mean: 24.8	

* This term denotes the teacher's conscious attempt to distribute his attention more equitably among his pupils.

† Based on one period of seventy-five minutes.

‡ These two pupils worked during all three periods in the craft shop on a project and were not in the classroom.

utes of the controlled sessions, F amassed two hundred and twenty-four contacts out of a grand total of one thousand, seven hundred and twenty-nine; this represents approximately twelve per cent of the total recorded contacts in the class. She and the pupil who amassed the next highest number of contacts, i.e., one hundred and seventy-four, together account for twenty-three per cent of the one thousand, seven hundred and twenty-nine contacts. F had slightly more than, and D slightly less than twice the mean number of contacts for the class. The combined total contacts of five pupils used for comparative purposes in the analysis of the "uncontrolled" sessions, i.e., A, Q, R, X and Y account for a total that is one and a half times as large as F's total contacts and double D's total. That is, the two pupils with the largest number of contacts did not take as large a relative share of the teacher's attention as they had taken in the "uncontrolled" situation. It must be noted, incidentally, that F received 37.6 contacts per hour in the first series and 72.6 in the second series—almost twice as many—after the teacher tried to even things up!

On the basis of the number of contacts registered in the three controlled sessions representing a total elapsed time of one hundred and eighty-five minutes, it is possible to compute the number of contacts per hour for each student. The mean number of contacts per hour in the uncontrolled sessions was 15.8 (see Table I); the mean number of contacts per hour in the controlled is 24.8 (see Table II).

CONCLUSIONS

Figure 1 is a columnar representation of the percentage of the total number of contacts with the teacher made by each individual² during eight "uncontrolled" class sessions totalling five hundred and fifty-five minutes and three "controlled" sessions totalling one hundred and eighty-five minutes.

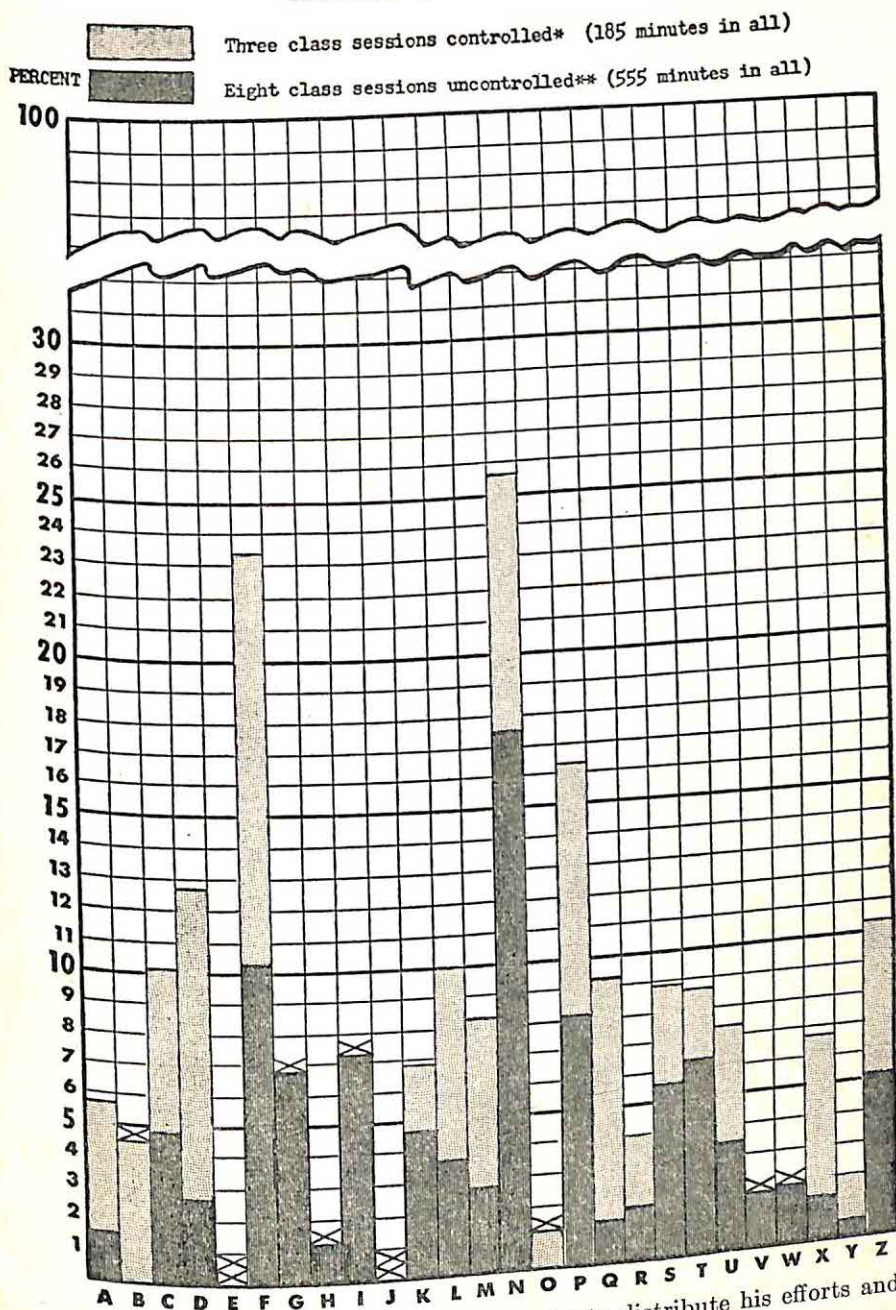
An examination of the cumulative histogram (Figure 1) emphasizes the relative changes in the distribution of the teacher's contacts within a class as a result of his deliberate attempts to ensure a more balanced distribution of his time and efforts among its members. The increase of nine (fifty-seven per cent) in the mean number of contacts per pupil per hour in the controlled sessions (Table II) over the mean number in the uncontrolled sessions (Table I) further underlines the change that occurred.³ However, despite the appreciable redistribution of teacher-pupil contacts in the situation, it seems reasonable to assume that the desideratum as regards the distribution of contacts with each pupil by the teacher was not realized.

It appears that we may not readily assume that even a teacher who displays a high degree of social sensitivity and who develops considerable rapport with his pupils will distribute his attention in the way which he and others on the basis of objective evidence and assessment of each pupil's needs would deem desirable and necessary. Likewise, it seems that once the teacher has been apprised of the "imbalance" of the distribution of his time and efforts among his charges, it is no simple task to redistribute his attention. None-

² Four of the pupils each missed one class period and are, therefore, not included in the tabulation.

³ Pearson's coefficient of variation, i.e., $V = \frac{6}{M} \times 100$ when computed on the basis of the data drawn from both the uncontrolled and the controlled sessions gives a V of 80.7 in the uncontrolled and a V of 57.5 in the controlled.

FIGURE 1—PERCENTAGE OF TOTAL NUMBER OF CONTACTS WITH
TEACHER BY EACH PUPIL DURING
ELEVEN CLASS SESSIONS



* Controlled—Teacher deliberately sought to distribute his efforts and attention more equitably in his class.

** Uncontrolled—Teacher was unaware of the distribution of his efforts and attention in his class.

X in column denotes absence of pupil from classroom during series of sessions.

theless, it appears likely that through the use of small, cohesive, working groups (1) of learners who encompass among them the scholastic and social skills required for enhancing both group and individual productivity, a teacher may make considerable improvement in the direction of becoming more accessible to all his charges and hence a more effective facilitator of pupil learning.

REFERENCES

- 1) Herbert A. Thelen, "Group Dynamics in Instruction: A Principle of Least Group Size," *The School Review*, March, 1949.

A READING-CENTER APPROACH WITHIN THE CLASSROOM¹

JOSEPH JACKSON

Public Schools, Dearborn, Michigan

The uniqueness of reading-center materials and techniques as utilized within that environment often posits queries about the possibility of applying such procedures to regular classroom reading instruction. If the potentialities of reading-center techniques are considered from the standpoint of their contribution to reading in general, an analysis of some of the underlying characteristics of their procedures may possibly serve to stimulate the reading accomplishment of youngsters ordinarily uninfluenced by the reading-center experience. It is, likewise, possible that experimentation within the classroom may divulge adaptations and uses of materials, which have not previously been evaluated sufficiently to establish their full impact upon reading as it is commonly taught to the great majority of children. One may rightfully say that a more articulated combination of classroom and reading-center approach to reading problems generally can result from this study.

The reading center of the Dearborn Public Schools is organized about two facets of purpose somewhat unique in the attainment of reading goals. The first is concerned with the remedial aspects of reading for youngsters who are transported to the center for five half-days a week of concentrated reading help but who return to their regular classrooms in the afternoon. The center affords opportunities to apply diagnosis and remedial work to the needs of severely handicapped readers who, because of average or better than average abilities, can be expected to attain a much higher status in reading skills. The second and more important facet concerns the in-service training of teachers in the techniques and procedures of remedial reading as practiced at the center. Five teachers from the elementary grades are assigned to the reading center for a period of six weeks. In the morning these teachers work with two pupils each at the center. In the afternoon the teachers participate in case discussions and assemble as well as develop reading

¹ Presented before the reading section of AERA, Cleveland, April, 1955.

materials which they utilize upon return to the classrooms. The present collection of materials contains one thousand thirty-four reading booklets—fourteen primers, forty-nine grade-one and one hundred twenty-one grade-two readers, and one hundred ninety-seven, two hundred ninety-five, two hundred nine and one hundred forty-nine booklets, respectively, for grades three to six, the increased number of the latter being explained in part by the increased skills within these grades. The dominant purpose of the reading center is to provide a remedial internship for teachers on the assumption that this practice will spread the techniques of teaching reading throughout the system and will materially reduce the number of reading problems within the classroom.

At the time that the local reading center was established in 1949, a national survey of reading remediation services was floated (1). The replies from nineteen public school systems revealed a pattern of function which Dearborn equaled or surpassed in such associated services as the training of teachers, reading diagnosis, the construction of reading matter, psychological case work, related consultation, and follow-up. The enrollment at the center is ten pupils. Since 1949 approximately one-hundred fifty teachers have experienced a six-week internship at the center. This study attempts to provide an insight into the possible classroom adaptation of the reading center methods as they might be applied to a cross section of the reading population.

THE PROBLEM

The reading center, as part of its remedial technique, stresses an individualized reading approach with its retarded readers. Once the reading level of a pupil has been determined, he is encouraged to read at his own rate and level provided that a standard of comprehension and of skill is maintained. Both story and skills types of reading materials are employed during the remedial process. An individual card is maintained, which serves at any time as an index of reading progress. If a negative history is found on this card, it indicates a need for the application of more specialized remedial techniques.

For the purposes of this study, the individualized reading approach was utilized as one experimental variable within the classroom and was substituted for the basic reading program. The classroom teacher's six-week internship had provided an under-

standing of the nature and application of the reading materials which she had helped to prepare. The individualized reading program as applied at the center is horizontal in scope, in that it is identified with a grade level of difficulty, as for example, grade three. To challenge the next level of reading-center materials, a student would have to be promoted to the next grade, as for example, grade four. Three hundred minutes are allotted the basic reading program weekly, the same as that for the individualized reading approach in the study.

While the horizontally individualized technique, as described, represents an extreme form of experimentation when compared to the conventional basic reader program, two additional experimental techniques were utilized which represented compromises between the basic and the individualized as practiced at the center. In one compromise approach, the use of reading-center techniques and of materials was restricted to approximately forty per cent of reading time, or to about one hundred and twenty minutes, the remaining one hundred and eighty being allotted to the basic reading program. In another compromise approach the application of reading-center materials was increased to seventy per cent of the reading time or to slightly more than two hundred minutes per week. The psychological influence of over-reading, namely over-learning, as imposed by the horizontal concentration of reading in the individualized approach would be less operative in these compromise approaches, should it exist, and would permit the students a varying amount of contact with the basic and co-basic readers. The teachers utilizing the compromise-experimental technique had likewise served a six-week internship at the reading center and were skilled in the application of reading-center techniques to classroom instruction.

The control counterpart to the experimental design was the conventional classroom reading situation organized about the basic and work study type of readers. Supplementation through the techniques and materials of the reading center was lacking. The teachers in the control aspect of the study have had no orientation in the use of reading center materials or of its techniques.

SUMMARY OF FINDINGS

Grade Three. The grade three groups in the study were equated upon such pre-test data as intelligence, achievement in reading,

and in the skill of map reading. The background data consisted of an initial evaluation, using the Metropolitan Primary II Reading Test and a final one using the elementary form of the Iowa Every Pupil Test, the change being made because of the marked growth of several pupils in reading ability. Two forms each of the Gates Basic Reading Tests and of the Dearborn Map Reading Test were likewise administered. The Monroe Basic Reading Test used for a summary appraisal of pupil progress in the local grade-three reading program was given as a terminal test only. In the remaining grades, however, the Iowa Every Pupil Tests were utilized for obtaining the comparative data of Table I, the choice of form being determined by the reading and grade status of each group. Seventeen youngsters were paired for the duration of the study. The average intelligence quotient of the conventional basic reading group was ninety-three, while that of the individualized was ninety-four. One can rightfully assume that the mental ability of several pupils within each group was below that of the reading-center group and hence involved that portion of the student body where reading problems might be common within the classroom. The grade three at the Maples School represented the control variable or the basic reading approach, while the William Ford School experimented with the free use of reading-center materials concentrated about a grade three level of difficulty. The reading center is located at the latter school.

A mean pre-test difference significant at the one per cent level of confidence favored the conventional reading group on the Gates pre-test, type C, understanding precise directions. In general, the pre-test differences on the Gates reading test, types A, B, and D, were likewise slightly advantageous to the basic reading group. The Gates reading tests represent a vital part of the diagnostic evaluation of the reading-center techniques.

The pre-test data of grade three as obtained by the Metropolitan reading test had indicated reading superiority of the conventional group. At the culmination of the experiment, the individualized reading group attained on the Iowa Every Pupil Test a status or grade level within a month of the conventional group. The comparisons in the succeeding grades were made through the Iowa tests only. Refer to Table I for mean score and grade level attainments of all groups. Since achievement on the Iowa tests is the preferred criterion of reading status within this grade locally, it appears that

TABLE I.—READING SCORES AND DIFFERENCE AFTER ONE YEAR USING VARIOUS METHODS OF READING INSTRUCTION

Grade 3B-A (N = 17)

	I.Q.	Pre-		Final	
		Score	G.L.	Score	G.L.
Conventional (Maples)	93	52.1	3.7	79.5	3.8
Individualized (Wm. Ford)	94	45.0	3.3	76.8	3.7
Months' difference			4		1

Grade 3B-A (N = 9)*

Compromise 40% (McDonald)	111	44.9	3.3	44.2	4.0
Individualized (Wm. Ford)	107	49.0	3.5	41.7	3.8
Months' difference			2		2

Grade 4B-A (N = 14)

Conventional (Maples)	112	53.2	4.2	59.9	5.1
Compromise 40% (Wm. Ford)	110	53.2	4.2	68.9	5.8
Months' difference			—		7

Grade 5B-A (N = 17)*

Conventional (Wm. Ford)	113	80.1	6.0	51.8	7.0
Compromise 70% (Maples)	114	78.2	5.8	58.9	7.6
Months' difference			2		6

Grade 6B-A (N = 17)

Conventional (Maples)	112	55.7	6.3	66.4	8.2
Compromise 40% (Wm. Ford)	112	55.2	6.3	59.7	7.7
Months' difference			—		5

Score differences $P > 0.05$

* Pretest-elementary form; final-advanced form.

the individualized reading content did not penalize the attainment of vocabulary or of comprehension when substituted for the basic reading program within grade three. As a rule the conventional group retained its original superiority on the Gates tests, types A, B, C, with type D showing a significant gain in the final evaluation. The conventional group, in addition, gained a statistically significant difference in the study skill, use of maps.

As an additional evaluation of the total experiment, Marion Monroe's Basic Reading Test ordinarily administered with the basic reader was given to the conventional as well as to the individualized reading groups of grade three. A statistically significant difference between means, 96 and 72 or 24, was established at the one per cent level of confidence in favor of the conventional reading group. This result is not surprising since the Monroe tests measure direct recall of the content of the basic reader and are used in conjunction with it. The mean classification category for the conventional group was *good* as compared to *poor* for the individualized approach.

The post-experimental follow-up of the achievements of the individualized and of the conventional groups a year later, or at the end of grade four, indicated the mean grade level status of each group to be 5.3 or likewise equal as they were at the termination of the experiment. Within grade four, the individualized group used in equal proportion the reading-center technique along with the basic reader. Through the follow-up data, one may infer that the experience with reading-center techniques and materials did not deter adjustment to the basal reader approach despite marked differences in the organization of content and procedure. One may, likewise, observe that the contributions of specialized reading techniques should be appraised upon residual improvement beyond the detail of the experiment lest one mitigate the influence of chance growth, teacher personality, teaching skill, class size, or time allotment as elements of remedial outcomes.

An attempt was likewise made in grade three to compare the achievement of a partially individualized basic reader group with that of a completely individualized one. The William Ford School participated again as the experimental variable, while the McDonald School acted as the control variable. However, the latter school limited the individualization to forty per cent or to one hundred and twenty of the three hundred minutes allotted weekly to read-

ing. The remaining time was devoted to the use of the basic and the co-basic readers.

The pre-test data indicated insignificant but accelerated abilities of the individualized group in reading, map reading and the Gates tests A, B, C with statistical significance for pre-test D, reading to denote details. The partially individualized or compromise group had slightly higher intelligence ratings. Only nine pupils remained for the two-semester period of study.

After a year's application of the two degrees of individualized control of reading content, the compromise group with forty per cent of time devoted to reading-center materials gained more in reading achievement than did the group utilizing the reading-center materials one hundred per cent of class time. On Gates tests C and D, reading to understand precise directions and reading to denote details, the group differences diminished to insignificant ones, major gains being made by the compromise group.

From the limited inference permitted from the above review, the problem needs to be more fully explored as to the degree to which reading materials contribute most critically in proportion to reading time within the regular classroom program. As suggested in grade three, the time appears to be less than one hundred per cent.

Grade Four. Within grade four, a conventional basic reading group was paired with a compromise group exposed forty per cent of the time to reading center materials of a grade four level of difficulty. The average intelligence quotient of the compromise group was 110 and that of the conventional was 112. Both groups attained identical pre-test scores on the Iowa Every Pupil Test of Reading and slight differences were likewise found for the Iowa Study Skills in such abilities as map reading, use of references, use of index, use of dictionary and alphabetization. Since work-study type of readers are part of the local basic reading program, it was desired to appraise possible losses or gains in such achievement as they may have been influenced by the compromise technique of the reading center which also incorporates skills into its individualized content. The pre-test data indicated insignificant group differences in comparative scores on tests A, B, and D of the Gates reading series and a significant difference on the side of the compromise group for test C, understanding precise directions.

If one again uses the final results of the Iowa Every Pupil Test

of Reading as the criterion of growth, larger if not significant gains were made by the compromise reading group. The interim gain is shown in Table I for grades four to six based upon the Iowa tests. This gain when interpreted through grade-level comparisons is of a seven-month difference between mean achievements. The compromise or experimental group achievement on the Iowa Every Pupil Test of Basic Skills, namely, map reading, use of dictionary and alphabetization had shown greater increases for the compromise group. The final tests of the Gates reading series portrayed the retention of superiority of the compromise group on the four types of skills evaluated. Fourteen students were paired within the grade four groups.

The grade four results permit one to generalize that the compromise group, oriented forty per cent of time through horizontally controlled content, did show larger reading gains than did the group exposed to the basic reader. The horizontal technique also failed to impose a penalty on the acquisition of study skills since, speaking conservatively, the attainments were at least equal.

Grade Five. Within grade five, still a different degree of individualization was utilized. In this part of the study, the reading center method was applied seventy per cent of time, the remaining time or ninety minutes weekly was allotted to the basic reading technique. The seventeen pupils equated for the study disclosed homogeneity in intelligence, reading achievement, and the work-study skills. The Gates reading pre-test scores indicated a significant superiority of the compromise group for test type A, appreciating general significance and test D, reading to denote details.

Here again the compromise group had shown a greater gain in vocabulary and comprehension through the Iowa Every Pupil Test of reading achievement. At the conclusion the compromise group attained a higher grade level status in vocabulary and in comprehension achievement. Because of the reading ability of the groups, the advanced edition of the Iowa Every Pupil Test was administered in the final evaluation instead of the elementary, thereby explaining the difference in the scores and corresponding grade levels. The compromise group likewise retained its relative position in achievement on four of five study skills tests. This was likewise true in the final evaluation of Gates Basic Reading Tests in which three of four outcomes maintained increments equal or above that of the pre-test status.

Grade Six. The experiment within grade six reverted to a forty per cent utilization of reading-center materials. The average intelligence quotient of each group was 112; the reading abilities at the beginning of the study were likewise identical. The conventional group was significantly superior at the start in the work study skill, use of maps and accelerated in the study skills use of index, use of dictionary and the reading of graphs, charts, and tables. The initial results of the Gates Basic Reading Tests also indicated a superior background on all four reading skills.

At the end of the experiment a year later, the conventional basic reading group showed superiority in silent reading and maintained its advantage on all five work-study skills. The conventional reading group likewise added to its differential of scores originally obtained through the Gates Basic Reading Tests, significantly so on type D, reading to denote details.

An additional variant was accepted in the grade six experiment which was lacking in the preceding grades. A substitute teacher was involved in teaching the experimental group for five weeks, while the regularly assigned teacher was concluding his internship at the reading center. Without this concession, the experiment would not have been possible at the grade six level. Since the substitution had taken place at the beginning of the semester, one might infer that the differences in results had been but mildly affected by the teacher changeover. Greater experience in the teaching of reading, however, must be recognized on behalf of the teacher of the conventional group. It was felt that this difference represented a major departure in this ability only within grade six.

OBSERVATIONS

(1) In three of four experimental grade situations, sufficient gains had been made in silent reading abilities of vocabulary and of comprehension to recognize that the application of reading-center techniques in toto or in part is worthy of consideration within the classroom.

(2) It is possible that reading gains made through the experimental use of reading-center materials may, in part, be attributed to the motivation inherent in the technique since recognition, the feeling of attainment, and the fulfillment of task are more discernible in the method of the reading center. The grade three teacher of the experimental group commented as follows: "The children's

interest is still keeping up. Time and again when told to put their books away, they asked for extra time."

(3) The children expressed their viewpoints in this way: "The stories in my book are interesting." "When I get a new series, I read more books." "I like to read because we don't have to read the same stories." "We don't have to sit in a circle in front of the room."

(4) Since pupils with sub-average mental capacities are least considered as candidates for reading-center or reading-clinic aid, special study should be given the potentialities of individualization for meeting the needs of slow learners especially those with intelligence quotients below 90.

(5) In order to rule out the booster effect of specialized reading techniques, of the teacher variable, and of the appraisal error, long-term follow-up should be instituted to more adequately disclose residual gains in reading comprehension along with the relative values of succeeding techniques.

(6) The Gates tests of basic reading were introduced into the study since they serve as criteria of growth at the reading center. It is questionable whether emphases per se were placed upon such outcomes as appreciating general significance, reading to predict outcomes, understanding precise directions, and reading to denote details, or whether such outcomes are a chance result of growth in the skill of reading comprehension. The inconsistent pattern of achievement of the control and of the experimental groups on the Gates tests seems to suggest such an inference. Teachers challenged the reliability of low scores about a broad sampling of reading content extrapolated, in part, from a grade level range of 2.8 to 12.5. Teachers of the experimental groups skilled in the application of the Gates test failed to gain advantages beyond those of a chance nature.

REFERENCE

- 1) Joseph Jackson, "Reading Remediation Services of Nineteen Public Schools." Bulletin No. 360, Department of Research, Public Schools, Dearborn, Michigan, pp. 19.

DIRECTED VERSUS INDEPENDENT DISCOVERY OF ESTABLISHED RELATIONS

ROBERT C. CRAIG¹

State College of Washington, School of Education

Experimental evidence favors some degree of learner activity over the specification of correct answers by the teacher or experimenter (3, 6, 7). But the results have been interpreted too broadly as implying that the more freedom of activity and, hence, the less external direction, the better (3, p. 15; 5, p. 141; 8, p. 80).

A recent experiment, similar in materials and design to the one reported here, dealt with the effect of external direction to aid the learner in his search for established relations among given concepts upon the learning and application of these relations (1). The results suggested that learner activity should be directed for efficient learning of principles. The advantage of direction, however, may be temporary. Advocates of unaided discovery may admit that fewer relations were learned when external direction was not provided, but they might well ask if those that were learned independently would not be remembered better and used longer than those learned with the assistance of the teacher or experimenter. Also, evidence is needed as to the effect of external direction on the learner's ability to solve future problems involving unlearned principles when external direction is not available. The hypothesis tested in the present experiment was that increased direction of discovery activity effects increases in learning without accompanying losses in retention or transfer.

MATERIALS AND PROCEDURES

Two groups of college students were each given a different amount of direction to help them discover the bases for the solution of a series of learning situations. A pre-test was used to measure initial knowledge of these relations. The number of relations actually learned by each subject was the difference between the number known at the final learning trial and the number known before training. The pre-test was given a second time as a post-test

¹ Now at American Institute for Research, Pittsburgh, Pa.

to measure the relative effects of the two degrees of direction on retention. A second post-test measured the post-training ability of each group to discover the bases for the solution of new situations independently. Confirmation of each learner's knowledge of post-test relations was obtained by asking him to write out the reasons determining his answers. The remaining paragraphs of this section describe the elements of the experimental materials and procedures in detail.

SUBJECTS

The subjects for this experiment were two groups of fifty-three sophomores and juniors enrolled in the experimenter's education courses at the State College of Washington. Forty-nine were men, fifty-seven were women. Each group was a fifty per cent random sample of the author's undergraduate students during the fall semester of 1953-54.

The two groups were similar with respect to initial test performance. The mean ACE score of the group receiving more direction was 99.24; S.D., 21.73. The mean ACE score of the group receiving less direction was 101.08; S.D., 22.48. The mean number of relations identified by the directed group on a pre-test composed of items organized on the same bases as the learning series was 6.00; S.D., 2.62. The mean for the second group on this pre-test was 5.52; S.D., 2.44.

PRE- AND POST-TESTS

The pre-test and first post-test for knowledge of relations consisted of sixty items—a group of four for each of fifteen organizational bases. A second post-test, consisting of forty items—a group of four for each of ten other organizational bases—was used to measure post-training ability to discover and use unlearned relations.

The following item illustrates the type of verbal problem used throughout this experiment:

(A) coin (B) plate (C) button (D) ball (E) wheel

The five words of each such item were selected so that the four are related in some manner. The learner's task is to select the word that does not belong with the other four. To solve this and similar

items efficiently, the learner must apply the general relation to the specific objects named.

The learner's determination of the basis for a correct response, including both the general formulation of the relation and its application to specific items is called "discovery of established relations." Spearman elaborated the process as one of his principles of cognition termed "the eduction of relations" (4, p. 63). His statement of the principle is, "The mentally presenting of any two or more characters (simple or complex) tends to evoke immediately a knowing of relation between them." Any one test situation is more restrictive than the general principle as stated, however, in that the relation of interest is a particular pre-selected or "established" one rather than any known. Formal learning is frequently so restricted. Of the many relations that a learner may perceive among the elements presented in a science lesson, for example, only the one selected by the teacher or curriculum maker will be considered important or, as an answer, correct.

Pre- and post-test items were based on relationships among the five words of each item. In each item the correct answer was determined by the sound of the words, the spelling of the words, familiar combinations of words, the meaning of the words, or combinations of sound, spelling, arrangement and meaning. The organizational bases and items have been published elsewhere (1). The example for sound relationships given below is illustrative.

Sound Relationships

Form two pairs of words beginning with the same sound.

1. (A) *cycle* (B) *seldom* (C) *sawdust* (D) *sausage* (E) *cellar*
2. (A) *rapping* (B) *wreckage* (C) *wrapper* (D) *reckon* (E) *reason*

Four items illustrating each relation were grouped together and separated from other items by spacing. No statements or other clues to the bases for correct response were included. The item groups were arranged in random order. If a subject solved all four of the items organized on a given basis correctly, he was given credit for knowing that relation. If he failed to solve any one of the four, he received no credit.

The tests were designed for group administration with all necessary directions at the top of the first page. The directions called the learner's attention to the existence of reasons for correct an-

swers but gave no clues to the nature of the reasons other than that provided by the grouping of items. These directions follow:

Words That Don't Belong

Directions: Each item of this test consists of five English words. In each item there is one word that does NOT belong with the others for some reason.

The items have been grouped in sets of four organized on the same basis, that is, the reason one word does not belong with the other words of the item is the same for each of the four items in a group. Different principles are used for different groups.

For each item, underline the word that does not belong with the others of the set. Answer every item. If you do not know the answer, guess.

Data from an independent but comparable group of one hundred and ninety-three students were used to determine test reliability. Since all subjects completed the tests within the time limits, split-half procedures were appropriate. As there were four items organized on each basis, two equivalent halves were formed by randomly choosing two of the four for each half-test. The two half-tests were scored separately for number of relations known, and the two sets of scores were correlated. The resulting correlations gave 0.93 as the estimated reliability coefficient for the pre-test and longer post-test and 0.90 for the forty-item post-test.

LEARNING MATERIALS AND TREATMENT

Groups of items of the same type that were used for pre- and post-tests were used for learning material. Two combinations of clues to the reasons determining correct responses to these items provided two different degrees of direction during discovery. The two combinations of clues were incorporated into two sets of learning materials suitable for group administration. The administration of the two sets of materials, each to a different group of subjects, constituted two different treatments during learning. A consistent procedure was used that had been thoroughly pre-tested in previous experimentation (1).

The learning materials for the group who searched for principles with relatively little direction (the independent group) consisted of items arranged in the same manner as the items of the pre-test; that is, items organized on the same basis were grouped together and separated from other items by spacing.

Each subject was immediately informed of the correctness of

right choices and of the incorrectness of wrong choices. Learners were instructed to continue to choose on each item until the correct response was discovered. A cardboard-backed answer sheet was used to inform subjects as to the correctness of their responses. When the subject marked the space corresponding to the correct answer choice, his pencil tip dropped into a hole in the cardboard backing. The holes corresponding to the right answers penetrated only one layer of the corrugated backing; thus, no holes were visible to the subject. If he blackens a space he believes to be correct and his pencil tip does not fall into a hole, he knows his choice is incorrect; if his pencil tip does fall into a hole he knows he has succeeded.

No information was given the independent group that might suggest the nature of the reasons for correct answers, but learners were told that the correct responses to all items of a given group were determined by the same organizational principle. With this information, the learner's attention is directed to similarities among the items of each group; he may search for this similarity to help him in his discovery. The instructions given this group were the same except that the following was added:

If your choice is correct, your pencil tip will pierce the answer sheet and fall into a hole in the cardboard backing. Thus you will know whether your choice is correct.

If your first choice is not correct, choose again. If the second space is also incorrect, choose again. Continue to choose until you discover the correct answer.

Do not spend too much time on any one item. Mark your choices on the separate answer sheet. Make no marks on the test booklet.

The more incorrect choices you make before discovering the correct answer, the lower your score will be. So, mark no more spaces than necessary to discover the answer. Do not erase.

The learning materials for the directed group differed from those for the independent group in that the learner was provided with a short general statement of the relationship common to the items of each grouping to direct his discovery. Each such statement was printed on the learning materials immediately preceding the item group to which it applied. The example which was presented to illustrate the several types of relationships also illustrates the nature of these statements and the placement of the statements with respect to the item groups.

The instructions given this group were identical with those for

the independent group with one important exception. The following sentence, "The principle applying to each group of four items is stated just above the group." was substituted for the sentence, "Different principles are used for different groups."

The learning treatments began three days after the administration of the pre-test. The sequence followed by each group of learners was as follows:

First Day: Discovery of correct responses to two illustrative items for each of five different organizational bases selected at random from the fifteen relations of the pre-test (Set A).

Third Day: Discovery of correct responses to two additional illustrative items for each relation of Set A.

Fifth Day: Discovery of correct responses to four illustrative items for each relation of Set A. Solution of all four items organized on a given basis without error was taken as proof of the knowledge of that relation.

Fifteenth, Seventeenth, and Nineteenth Days: Repetition of the sequence for the first, third, and fifth days with a second group of five randomly selected relations (Set B).

Twenty-ninth, Thirty-first, and Thirty-third Days: Repetition of the sequence for the first, third, and fifth days with the remaining five relations of the pre-test (Set C).

Both post-tests were administered on the third day following the end of the learning sequence or three days after the final learning experience with Set C, seventeen days after the final learning experience with Set B, and thirty-one days after the final learning experience with Set A.

ANALYSIS OF RESULTS

The items of the pre-test and those of the learning series were organized on the same fifteen bases. In this section, performance on the pre-test and during learning are analyzed first to identify the reasons for correct responses which were actually learned during the experiment. Then learners' responses on the post-training administration of the pre-test and subjects' reasons for their responses to this test are studied to reveal the number of relations retained. Finally, learners' responses to a second post-test and their reasons for these responses are analyzed to compare the post-training ability of the two groups to discover and use unlearned relations.

TABLE I.—NUMBER OF RELATIONS LEARNED BY DIRECTED AND INDEPENDENT GROUPS

(Group Means and Variances for Each Set)

Set of Principles	Statistic	Directed Group	Independent Group	Significance Tests*
A	Mean	2.280	1.520	(<i>t</i>) 3.84
	Variance	0.910	1.176	(<i>F</i>) 1.29
B	Mean	2.560	1.520	(<i>t</i>) 5.84
	Variance	0.923	.760	(<i>F</i>) 1.21
C	Mean	2.560	1.920	(<i>t</i>) 3.03
	Variance	1.257	1.110	(<i>F</i>) 1.09
All sets	Mean	7.400	5.080	(<i>t</i>) 5.10
	Variance	4.667	6.327	(<i>F</i>) 1.36

* Students' *t*'s all significant at 0.01 level; all *F*'s not significant at 0.05 level.

NUMBER OF RELATIONS LEARNED

The pre-test consisted of fifteen groups of items—a group of four for each of fifteen bases—the same fifteen used for learning. A subject who solved all four of the items organized on a given basis on the pre-test or during learning demonstrated his knowledge of the relation at that point in the experiment. The number of relations actually learned by each subject was obtained by determining the number known after learning that were not known prior to learning. The mean number learned and the variance in the number learned by each experimental group are shown in Table I.

Differences in the mean number of organizational bases learned consistently favored the directed group. Students' *t* corresponding to the mean differences for each set was statistically significant at the 0.01 level of confidence. *F*-ratios of the variances for each set were not statistically significant. Directed discovery resulted in the learning of more relations than independent discovery.

NUMBER OF RELATIONS RETAINED

The test administered to measure retention was administered three days after the final learning trial with Set C. This was seven-teen days after the final trial with Set B and thirty-one days after

TABLE II.—PERCENTAGES OF LEARNED RELATIONS RETAINED BY GROUPS AT THREE INTERVALS AFTER LEARNING

AT THREE INTERVALS AFTER LEARNING								
Group	N*	Mean Per Cent Retained	No. of Subjects by Per Cent Retained					Chi-square
			00.0	20.0 25.0	33.3 50.0 66.7	75.0 80.0	100.0	
Three days after learning (Set C)								
Directed	51	92.61	0	0	4	9	38	4.267†
Independent	46	90.61	2	0	2†	2	40	
Seventeen days after learning (Set B)								
Directed	53	79.57	4	0	13	4	32†	2.256†
Independent	47	75.47	6	2†	9	0	30	
Thirty-one days after learning (Set A)								
Directed	53	80.49	0	0	19	2	32†	15.392§
Independent	47	62.32	12	0	13	0	22	

* Number in group who learned at least one relation.

† The bar connects frequencies that were combined for the chi-square computation.

‡ Not significant at the 0.05 level.

§ Significant at the 0.001 level.

the final trial with Set A. Thus, the retention of the two groups was compared over three different intervals of time.

Table II shows the number of subjects of each group retaining each percentage of the learned relations of each set. Because there were only five in each set, only the percentages indicated in the frequency distribution in Table II could occur.

Inspection of Table II indicates little difference among the two groups with respect to retention of Set C, three days after learning, or Set B, seventeen days after learning. With Set A, however, it seemed that more of the directed group than the independent group retained no relations. The mean percentage retained by the independent group continued to decline as the interval since learning increased from seventeen days to thirty-one days; that for the directed group did not. Chi-squares corresponding to the frequency distribution of the two groups for Set C and B were not significant, but the chi-square for the frequencies of Set A was significant at

the 0.001 level. After approximately one month, the directed group retained more learned relations than the independent group.

Analysis of learners' reasons for each response of this test confirmed their knowledge of the organizational bases for item groups solved correctly and their lack of knowledge concerning the bases of items failed. Statements classified as incorrect were ambiguous, admitted guesses or relationships based on faulty observation and analysis of the words. Careful pre-selection of items and the requirement that four items be solved without error before credit was given for a relation apparently eliminated correct reasons other than those anticipated by the experimenter. Examples of the types of statements considered correct and incorrect are available in a previous publication (1).

POST-TRAINING DISCOVERY OF NEW PRINCIPLES

A second post-test composed of items based on relations entirely new to the learner was administered three days after the final learning trials. Table III summarizes the results of this test by giving the mean number discovered by each group.

Each group appeared to improve in its ability to discover relations for it solved a greater proportion of these post-test items than of the pre-test items prior to training; and a supplementary analysis indicated that the post-test was more difficult than the pre-test.

The supplementary analysis utilized the independent group of one hundred and ninety-three students of the reliability determination. For this independent group the difficulty of the pre-test relations, in terms of the percentage of the group who answered the corresponding items correctly without training, ranged from two

TABLE III.—NUMBER OF POST-TEST RELATIONS
DISCOVERED INDEPENDENTLY
(Group Means and Variances)

Group	Mean*	Variance†
Directed	4.880	2.193
Independent	5.000	1.667

* Student's *t* for difference in means is 0.445, not significant at 0.05 level of confidence.

† *F*-ratio of variances is 1.315, not significant at 0.05 level of confidence.

to eighty-two per cent, with a mean of 41.7 per cent. The range of difficulty for the post-test of new relations was also two to eighty-two percent, but the mean was 35.1 per cent.

Differences between the two groups in mean number of new relations discovered are small and not statistically significant. Again, subjects' reasons for responses confirmed the number of organizational bases known by each group.

DISCUSSION

The results of the present experiment affirm Thorndike's conclusion that "... the widespread limitation of guidance ... to designating errors is a sign of weakness in the technique of teaching" (9, p. 147). These college students profited from greater guidance or direction in that they learned more relations. This finding is in sharp contrast with results obtained when correct answers are specified for learners in advance. Stacey, with learning materials similar to those used in the present experiment, found specification of correct answers in advance of pupil discovery activities to be inferior to merely "designating errors" (6). On the other hand, the present results agree with previous experimentation as to the effect of direction during discovery on learning (1). A reasonable interpretation of these apparently conflicting results seems to be that subjects use and benefit from help given them in their search for bases determining correct responses; but, when correct responses are specified in advance of activity, they neither search nor discover. It is probable that they just try to remember specific responses.

The present investigation went beyond previous experimentation in that the effect of increased direction upon retention and transfer was investigated. The hypothesis that increased external direction that helps learners in their search for established relations effects an increase in the number learned without an accompanying loss in retention or transfer was confirmed for the present context. The group receiving the greater direction retained those learned relations as well as the group learning by more independent discovery. Further, the results suggest that as the interval of time after learning increases to about one month, the group receiving more direction retains a greater proportion of what was learned than the other group. This suggestion has since been confirmed with other subjects and mathematical material (2). If there were no differences

in the proportion of learned relations retained, the directed group would still have an advantage because they learned more and, hence, retained more in terms of number.

No differences were found in the transfer problem of solving items involving unlearned relations. Both groups were found to improve in their ability to solve this type of item as a result of the learning experience.

The results of the present experiment add to the evidence indicating that teachers or experimenters should be liberal with suggestions to aid discovery of principles. The desirable degree of direction would include general statements of the relations to be learned.

The limited scope of this experimentation is recognized. The available evidence needs confirmation with other materials, other subjects, other training and retention intervals. Further information concerning the effect of discovery experiences with varying amounts of direction over periods of months or years is critical in the determination of effects on future discovery.

Many have advocated relatively independent problem-solving in the belief that learning situations should be similar to anticipated transfer situations. This point of view rests on the assumption that future discovery of principles will be through independent problem-solving, hence, more like pupil self-discovery than directed discovery. A different view is that problem-solving and discovery are never independent except in the sense that no one is physically present to prompt the learner. Principles previously learned in an area serve to direct discovery. Out-of-school discovery is not independent but directed by the knowledge gained under the direction of previous teachers. The more direction of this kind available to the learner, the more effective his discovery of new relations. The cumulative effect of greater learning through directed discovery over months or years may offset the effect of any lack of similarity between learning and transfer situations and prove to be the best preparation for new discoveries.

SUMMARY

An attempt was made to determine the effect of directing learners' discovery of established relations upon retention and the ability to discover new relations.

Each of two groups of college students was given a different

amount of direction during discovery of the bases determining solution of multiple-choice verbal items. The group receiving the greater direction, including short summary statements of the bases determining correct responses, learned more relations in each of three trials. Three days after learning and seventeen days after learning, the two groups did not differ in the proportion of learned relations retained. After thirty-one days, the directed group retained a greater proportion of those learned than the more independent discovery group. After a total of thirty-three days of training, the two groups improved about equally in their ability to solve problems organized upon unlearned bases.

This evidence indicates that teachers and experimenters should be liberal with information designed to assist learners in the discovery of principles. Large amounts of external direction now may help to insure that the learner will have an adequate background of knowledge to direct his future discovery.

REFERENCES

- 1) R. C. Craig, *The Transfer Value of Guided Learning*. New York, Teachers College, Columbia University, Bureau of Publications, 1952.
- 2) R. C. Craig, *Verbal Direction and the Learning of Principles*. Paper read at Oregon and Washington Psychological Associations, Portland, May, 1955.
- 3) T. R. McConnell, "Discovery Versus Authoritarian Identification," in "The Learning of Children," *University of Iowa Studies in Education*, 9: 1-62, 1934.
- 4) C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*. London, Macmillan, 1923.
- 5) H. F. Spitzer, "Learning and Teaching Arithmetic," pp. 120-142 in National Society for the Study of Education, *The Teaching of Arithmetic*, Fiftieth yearbook, Part II. Chicago, University of Chicago Press, 1951.
- 6) Esther Swenson, G. L. Anderson, and C. L. Stacey, *Learning Theory in School Situations*. Minneapolis, University of Minnesota Press, 1949.
- 7) C. L. Thiele, *The Contribution of Generalization to the Learning of the Addition Facts*. New York, Teachers College, Columbia University, Bureau of Publications, 1938.
- 8) C. L. Thiele, "Arithmetic in the Middle Grades," pp. 76-102 in National Society for the Study of Education, *The Teaching of Arithmetic*, Fiftieth yearbook, Part II. Chicago, University of Chicago Press, 1951.
- 9) E. L. Thorndike, *The Psychology of Wants, Interests, and Attitudes*. New York, Appleton-Century, 1935.

EFFECTS OF NON-PROMOTION ON EDUCATIONAL ACHIEVEMENT IN THE ELEMENTARY SCHOOL

WILLIAM H. COFFIELD

Alabama Polytechnic Institute

AND

PAUL BLOMMERS

State University of Iowa

INTRODUCTION

Since 1904, when Superintendent W. H. Maxwell published his startling report on retardation in the New York City schools, the problem of liberal promotion standards vs. promotion based on rigid minimum standards of achievement has been subjected to much serious consideration by educators. Today, some half a century later, the issue remains unsettled. This is particularly true at the elementary school level where practices extend through various degrees of compromise from adherence to a rigid minimum standard of achievement to automatic annual promotion. During this period, a number of research reports pertaining to this problem have appeared in the literature. These include investigative studies of age-grade status, school policies and practices, the extent of non-promotion, the causes of grade repetition, and the effect of non-promotion on the pupil. In addition to such reports, hundreds of articles advocating an opinion or describing a local practice have also appeared.

It is clear that the trend during this period has been in the direction of a decrease in the relative frequency of failure in the form of non-promotion. Larson (10) reports that at the turn of this century some fifty per cent of the pupils had experienced failure by the time of completing their elementary schooling, that this proportion was approximately halved by the thirties, and that at the present (1954) time it is of the order of ten per cent. In no sense, however, should this trend be interpreted as indicative of a settled issue. Vituperative critics of the present-day school refer to it as "the only institution this side of heaven that rewards intention as generously as it does accomplishment" (6) and decry the practice of promoting

"all children at the end of each academic year regardless of whether their work has been good, bad or indifferent." (4) (See also 6, 12).

In a very general way, the major arguments for and against automatic annual promotion may be set down as follows:

Pro

- (1) A child must experience success if maximum development is to occur.
- (2) Failure or the fear of failure leads to frustration which results in a thwarting of development.
- (3) Individual differences are of such character as to make the imposition of a single minimum standard completely incompatible with the aim of universal elementary education.
- (4) There is very little, if any, evidence to show that ultimate mastery of school work is enhanced as a result of grade repetition.
- (5) A failed pupil represents an added operational cost which cannot be justified in terms of the negligible gain in mastery which accrues from repetition.
- (6) The imposition of minimum standards tends to increase pupil variability in the upper grades with respect to age, interests, and physical and social maturity, and accentuates the problem of dealing with individual differences at this level.

Con

- (1) If no minimum standards are maintained for promotion, children will lose respect for scholarship and the quality of their school work will degenerate.
- (2) Since most people are at one time or another in the course of their lives confronted with a failure situation of some type, school failure is consistent with reality and may serve to prepare certain pupils to face life more intelligently.
- (3) Automatic promotion offers no incentive to the brighter, hard-working child and appears to reward the dull and indolent.
- (4) The automatic promotion of a slow learner increases the degree to which he lags behind his classmates and accentuates the problem of dealing with individual differences at the upper grade levels.

Possibly some of the arguments on both sides are valid to some degree. Clearly, some are purely statements of opinion—others have some evidence in their support. Briefly, such evidence as exists indicates (1) that when failed and non-failed pupils are paired off or matched in terms of mental ability, there is little, if any, over-all difference in their educational achievement during the next one or two years (3, 7); (2) that when failed and non-failed pupils are matched in terms of mental ability, socio-economic status and age, there is at most a very small difference, if any, favoring the non-failed pupils in quality of personal and social adjustment (2, 8); (3) that when the pupils of classrooms threatened with failure at the outset of a semester are compared with those of classrooms assured promotion, there are no noticeable differences in over-all achievement at the end of this semester, though teacher opinion seems to be indicative of a somewhat higher pupil morale in the latter type of classroom (11); and (4) that when schools are classified as high and low rate of non-promotion schools, the average level of pupil achievement is higher in the latter type school while the variability in pupil achievement is approximately the same for both types (1, 5). If, as the present evidence seems to suggest, there is little apparent difference in either the subsequent educational achievement or the personal or social adjustment of matched failed and non-failed pupils, then determination of policy must rest on other bases than these.

The writers recently undertook an investigation intended to add to the accumulating evidence on the relative quality of educational achievement of matched promoted and non-promoted pupils. The purpose of this paper is to report the methodology and findings of this study.

THE SUBJECTS

Superintendents of three hundred and two Iowa school systems which had a long history of participation in the Iowa Basic Skills Testing Program were asked to identify from among pupils enrolled in grade seven in the 1953-54 school year those who had experienced failure once since, but not during or before second grade, and to indicate the grade in which this failure occurred.¹ Replies received from two hundred and eighty-nine (95.7 per cent) of these

¹ Pupils who experienced failure during or before second grade were not included because test results were not available for any grade below third.

superintendents resulted in the identification of one hundred and ninety pupils—including forty-three whose failures occurred the previous year in grade seven—for whom test results were available in the files of the Iowa Testing Programs.

The one hundred and forty-seven pupils who had experienced failure in one of the grades from three through six were further divided into two groups. The pupils (ninety-three) in one of these groups were each matched with a pupil who was a promoted classmate in the grade in which the failure occurred, on the basis of the particular achievement variable studied. These ninety-three pupils were all those among the one hundred and forty seven for whom such matching classmates could be found. The remaining pupils (fifty-four) comprising the other group were similarly matched with promoted pupils, but in this case the promoted matchees were selected from another school system. Hereafter, these two types of matching will be referred to as *internal* and *external*, respectively.

Ideally, of course, we are concerned with the comparison of a failed pupil's educational development as against his development had the failure not been imposed. Obviously, since a given pupil cannot at the same time be both promoted and failed, the use of a matching pupil is essential. In matching a failed pupil with one of his own schoolmates (internal), the general conditions under which the failed pupil would have studied had he been promoted are maintained for the matchee. However, it is to be expected that in this type of matching a bias favoring the promoted group will be introduced, for even though each pair is perfectly matched in terms of obtained test scores, the fact that one member is passed while the other is not indicates that in the best judgment of the room teacher, some difference in favor of the promoted pupil exists between them. It was for this reason that external matching was also employed. Here, obviously, the general conditions under which the failed pupil would have pursued his education had he been passed are not maintained for the promoted matchees. On the other hand, the possible bias introduced as a result of internal matching is not as likely to be present—at least not to the same degree.

THE CRITERION VARIABLES STUDIED

As has been indicated, the concern in the investigation here reported was with educational achievement. The following standard scores (grade equivalents) yielded by the Iowa Tests of Basic Skills (9) were employed as criterion measures:

- (1) Reading Comprehension—Test A1.
- (2) Work Study Skills—Composite for Section B.
- (3) Language Skills—Composite for Section C.
- (4) Arithmetic Skills—Composite for Section D.
- (5) An over-all composite based on the sum of the above four scores.

These five measures were available for each failed subject (1) for the year in which failure occurred; (2) for the year following failure; (3) for two years following failure—except for the forty-three pupils who incurred failure in grade seven; and (4) for the 1953-54 school year during which all failed pupils involved were enrolled in grade seven.

COMPARISONS STUDIED

The comparisons made were of two general types. In the first type (hereafter designated as Type 1) the failed pupil was, in effect, compared with himself, whereas in the second type (Type 2) the failed pupil was compared with his promoted matchee. Several comparisons involving different periods of elapsed time were made under each general type. These individual comparisons which were made with reference to each of the five criterion variables were as follows:

Comparison 1.1²: Scores reported for failed pupils for year of failure with scores reported for these same pupils one year later (i.e., during the second year in the same grade).³

Comparison 1.2: Scores reported for failed pupils for year of failure with scores reported for these same pupils two years later (i.e., when the pupils were in the next higher grade).

Comparison 2.1⁴: Scores reported for failed pupils for year fol-

² I.e., Comparison 1 under general Type 1.

³ The Iowa Tests of Basic Skills are typically administered in Iowa schools during the latter part of January. Thus the initial measurements were obtained after the pupil had spent approximately five-ninths of the school year in the grade failed. The final measurements in the case of Comparison 1.1 were obtained one year later and after the failed pupil had spent approximately five-ninths of the school year in repeating the grade failed. While the pairs of scores thus obtained are lower than would be the case had the measurements been taken at the close of the school year, it is doubtful that the comparative differences are much affected. In any case, the validity of the comparisons for the basic purpose at hand should not be seriously affected.

⁴ I.e., Comparison 1 under general Type 2.

lowing failure (i.e., second year in same grade) with scores reported for promoted matchees for this same year (matchees in next higher grade). Here lapsed time is the same for both members of a pair, but the grade placement differs by one full grade.

Comparison 2.2: Scores reported for failed pupils for second year following failure (i.e., next higher grade) with scores reported for promoted matchees for this same grade. Here the grade placement is the same for both members of a pair, but lapsed time differs by one full year.

Comparison 2.3: Scores reported for failed pupils with scores reported for promoted matchees when both were in grade seven. Here, again, the grade placement is the same for both, but lapsed time differs by one year.

Comparison 2.4: Scores reported for failed pupils in grade seven with scores reported for promoted matchees in grade eight. Here lapsed time is the same for both members of a pair, but the grade placement differs by one full grade.

STATISTICAL ANALYSIS

A very brief account of the statistical procedures employed in analyzing the data for the various comparisons studied is presented in this section.

Both Type 1 comparisons (1.1 and 1.2) were analyzed by application of a *t*-test of the significance of the difference between means of related samples. Separate tests were applied for each of the criterion variables by grade in which failure occurred.

The analysis employed with the Type 2 comparisons in the case of the four content area test scores differed from that used in the case of the over-all composite score. It will be recalled that each failed pupil was paired with a promoted pupil solely on the basis of performance on the particular content area test under investigation. Regardless of whether the matching is internal or external, comparisons based upon it may be expected to be biased in favor of the promoted pupils, for while the two pupils may be perfectly matched with regard to the specific area in question, it is reasonable to expect that the promoted pupils will be superior in other areas and, therefore, less retarded in general.⁵ To control as much as pos-

⁵ Ideally, failed and promoted pupils should have been paired on the basis of test battery profiles. Unfortunately, such pairs are virtually nonexistent.

sible this factor of general or over-all retardation, the specific area test scores for the pairs of failed and promoted pupils were analyzed by the method of analysis of covariance with the over-all composite score being used as the statistical control variable. Separate analyses were made for each content area test by grade in which failure occurred. All Type 2 comparisons involving the area test scores were analyzed in this manner.

Variation in general or over-all retardation between members of a pair was not a matter of concern in the case of the Type 2 comparisons involving the over-all composite as a criterion since in such cases the individuals were matched on the basis of their over-all composite scores. Hence, all Type 2 comparisons involving the over-all composite scores were analyzed by application of the *t*-test for related samples.

RESULTS: TYPE 1 COMPARISONS

The findings with respect to both Type 1 comparisons are presented in Table I. The numbers of pupils involved at the different grade levels varied slightly from test to test as well as from the one to the two-year comparisons. However, the approximate numbers of pupils in the grade groups were fifty-five, forty, twenty-five, twenty and forty for grades three through seven, respectively. As has been indicated, the individual pupil scores were expressed in terms of a grade-equivalents scale. The numbers comprising such a scale are two digit numbers with the first digit (i.e., the digit in the ten's place) representing the grade level and the second digit representing the number of months spent in this grade. Thus a grade-equivalent score of 43 represents a performance corresponding to that of a typical pupil at the end of three months in grade four. Grade-equivalent scores ending in zero represent typical performance at the beginning of a school year, while those ending in nine represent typical performance at the close of a school year.

The figures in Column 1 of Table I are the means of grade-equivalent scores made on the various criterion tests by failed pupils. Separate means are reported for groups of pupils organized according to the grade failed. In each case, the means shown in this column are based on measurements taken during the year in which failure occurred. Columns 2 and 3 present corresponding means for the same pupils, except that they are based upon measurements taken one and two years later, respectively. The one-and two-year

TABLE I.—GAINS MADE BY FAILED PUPILS ONE AND TWO YEARS AFTER FAILURE

Grade Failed	Test	Means at Time of Failure (Col. 1)	Means 1 Year Later (Col. 2)	Means 2 Years Later (Col. 3)	1 Year Gains (Col. 4)	2 Year Gains (Col. 5)
3	Reading	29	32	37	3	8
	Study skills	32	34	40	2	8
	Language	27	32	39	5	12
	Arithmetic	31	34	39	3	8
	Over-all	30	33	39	3	9
4	Reading	32	36	45	4	13
	Study skills	34	41	48	7	14
	Language	33	40	47	7	14
	Arithmetic	35	40	50	5	15
	Over-all	34	39	48	5	14
5	Reading	42	50	55	8	13
	Study skills	43	50	60	7	17
	Language	41	50	55	9	14
	Arithmetic	43	52	61	9	18
	Over-all	42	50	58	8	16
6	Reading	48	54	60	6	12
	Study skills	48	58	64	10	14
	Language	48	56	61	8	13
	Arithmetic	54	61	68	7	14
	Over-all	50	58	63	8	13
7	Reading	55	62		7	
	Study skills	61	63		2	
	Language	54	62		8	
	Arithmetic	62	69		7	
	Over-all	58	64		6	

gains shown in Columns 4 and 5 are simply the differences between the means based on measurements obtained during the year in which failure occurred and the corresponding means based on measurements made one (comparison 1.1) and two (comparison 1.2) years later. In a logically loose and quantitatively approximate sense, these gains may be viewed as corresponding to school progress in months or in years and months according to whether they are one or two digit numbers.

Since the tests were typically administered in late January, the

expected or average scores for grades three through seven respectively are 35, 45, 55, 65, and 75. During the year of non-promotion, then, the pupils failing third grade were on the average for these tests only slightly more than one-half year retarded, while those failing the fourth and fifth grades were retarded by slightly more than a year, and those failing sixth and seventh grades by more than a year and a half.

After the lapse of one year, during which these pupils remained in the same grade, statistically significant mean gains were achieved by all groups on all measures save the sixth grade group in reading and the seventh grade group in work-study skills. Over all grades and tests these gains averaged about six months and ranged from two months for the grade three group in work-study skills to one year for the grade six group in this same content area. In no case, however, were the gains sufficient to bring the one year later means up to the grade norm. In other words, these pupils were still retarded, even after repetition of the same grade.

When the tests were administered to these groups after a lapse of two years (the pupils were then in the next higher grade), the gains averaged approximately one year and three months and ranged from eight months to one year and eight months. Again, however, the gains were insufficient to bring any of the groups up to the grade norm on any of the criterion variables studied.

A word of caution should be introduced lest the inference be drawn from Table I that gains tend to be greater for pupils experiencing failure in the upper elementary grades. This may or may not be true, but the data presented in Table I cannot validly be cited as evidence, since two distinct test batteries—an elementary and an advanced—are involved. While these batteries have been carefully equated, the fact remains that any differences in mean gains from grade level to grade level could possibly be due to differences in scale.

RESULTS: TYPE 2 COMPARISONS

The findings regarding comparisons 2.1 and 2.2 are summarized in Table II. The values shown in this table are in terms of the same type of unit or scale as the gains reported in Table I (Col. 4 and 5). Except in the case of the over-all values reported, they represent differences between means of the grade-equivalent scores of groups of matched failed and promoted pupils after an analysis of covari-

ance adjustment to control differences in general or over-all retardation, the original matching having been done on the basis of the particular test variable in question. No such adjustment was needed for the over-all criterion. These differences were all found by subtracting the means for the failed groups from those of the corresponding promoted groups so that positive differences are in favor of the promoted group. As was true of the Type 1 comparisons, the numbers of pupils at the various grade levels varied slightly from test to test as well as from comparison to comparison. However, the approximate numbers of pupils in the grade groups three

TABLE II.—ADJUSTED DIFFERENCES BETWEEN MEANS OF INTERNALLY AND EXTERNALLY MATCHED FAILED AND PROMOTED PUPILS ONE YEAR AFTER FAILURE AND WHEN BOTH ARE IN NEXT HIGHER GRADE

Grade Failed	Test	Adj. Mean Diff.* after 1 Yr.		Adj. Mean. Diff.* at Next Grade	
		Internal Grp.	External Grp.	Internal Grp.	External Grp.
3	Reading	9†	3	5†	-2
	Study skills	4†	4	-2†	-4
	Language	6†	4	-1	-5
	Arithmetic	5†	6†	1	-1
	Over-all‡	6†	5†	1	-3
4	Reading	6	10†	-1	2
	Study skills	6†	2	-1	-8†
	Language	6†	11†	-3	-1
	Arithmetic	0	5†	-5†	-8†
	Over-all‡	3	6†	-3	-6†
5	Reading	9	0	6	-4
	Study skills	4	5	-9†	-7
	Language	8	3	-2	-5
	Arithmetic	3	6	-3	-7
	Over-all‡	3	3	-4	-6†
6	Reading	12	-3	-5	-6
	Study skills	9	4	4	-1
	Language	-2	7†	-3	2
	Arithmetic	6	4†	-1	0
	Over-all‡	9	2	0	-4

* Positive differences favor promoted groups.

† Difference statistically significant beyond five per cent level.

‡ Not adjusted by analysis of covariance.

through six were forty, twenty, eighteen, and seven, respectively, for the internally matched pairs, and fifteen, fifteen, six, and fifteen for the externally matched pairs.

The first two columns of differences were based on measures taken one year after the failures were imposed (comparison 2.1), that is, after both groups had experienced an additional year of schooling during which the failed pupils remained in the same grade and the promoted pupils advanced to the next higher grade. Regardless of type of matching (internal or external) these differences tended to favor the promoted pupils, though statistical significance obtained in only about one-third of the situations. Considering all comparisons, these differences averaged about six months in the case of the internally matched pairs and about four months in the case of the externally matched pairs.

The last two columns of differences in Table II were based on measures taken when both groups were in the next higher grade (comparison 2.2), that is, after the failed group had experienced two additional years of schooling as compared with one for the promoted group. As would be expected, these differences tend to favor the failed group. Even so, relatively few of them (less than one-fifth) proved to be statistically significant. Considering all comparisons, these differences averaged only about one month in the case of the internally matched pairs and about three months in the case of the externally matched pairs.

The findings regarding comparisons 2.3 and 2.4 are summarized in Table III. The design of this table and the character of the values reported in it are like those of Table II. The first two columns of differences were based on measures taken when both groups were in grade seven (comparison 2.3), that is, had reached or were near the end of their elementary schooling, depending upon the type of school organization in effect. It is important to recognize that in this comparison the failed groups involved have spent one more year in school than the promoted groups. Only three significant differences were found and these favored the promoted group. Considered as a whole, the differences observed are virtually negligible regardless of whether the matching is internal or external. These differences average only two months in the case of the internally matched group and zero months in the case of the externally matched groups. Thus, the evidence indicates that the additional year spent in school by the failed pupils had not by the close

TABLE III.—ADJUSTED DIFFERENCES BETWEEN MEANS OF INTERNALLY AND EXTERNALLY MATCHED FAILED AND PROMOTED PUPILS WHEN BOTH ARE IN GRADE SEVEN AND WHEN THE FAILED PUPILS ARE IN GRADE SEVEN AND THE PROMOTED PUPILS IN GRADE EIGHT

Grade Failed	Test	Adj. Mean Diff.* in Grade 7		Adj. Mean Diff.* in Grades 7 vs. 8	
		Internal Grp.	External Grp.	Internal Grp.	External Grp.
3	Reading	8	-10	19†	5
	Study skills	10†	9	15†	16
	Language	9†	0	14†	10
	Arithmetic	1	1	8†	8
	Over-all†	5†	-1	13†	8
4	Reading	2	10	8	18†
	Study skills	9	-3	6	5
	Language	4	4	7†	13†
	Arithmetic	-2	0	3	6
	Over-all†	0	1	5	10†
5	Reading	-1	2	0	18†
	Study skills	1	3	5	9
	Language	-1	-5	4	14
	Arithmetic	2	-1	11†	8
	Over-all†	-3	-3	0	10
6	Reading	-5	-6	5	3
	Study skills	4	-1	14	2
	Language	-3	2	8	3
	Arithmetic	-1	0	8	2
	Over-all†	0	-4	9	0

* Positive differences favor promoted groups.

† Difference statistically significant beyond five per cent level.

‡ Not adjusted by analysis of covariance.

of their elementary school training enabled them to surpass in achievement their regularly promoted counterparts.

The differences reported in the last two columns of Table III were based on pupil performances measured when the failed groups were in grade seven and the promoted groups in grade eight (comparison 2.4), that is, when the number of years spent in school since failure was the same for both groups. In this case, regardless of whether matching is internal or external, all differences favored the promoted pupils. However, only about one-fourth of these dif-

ferences were found to be statistically significant. Considering all comparisons, these differences averaged close to a school year (eight months) for both internally and externally matched groups.

RESULTS: SUBSIDIARY PROBLEMS

As by-products of the data collected for this study it was possible to investigate four subsidiary problems.

The first of these had to do with the effect of the grade level at which failure occurred. Using only the five criterion scores obtained for the failed pupils when they were in seventh grade, analyses of variance were effected to test the hypothesis that the level of seventh grade achievement was the same for a group experiencing failure in one grade as for a group experiencing failure in any other grade. The evidence did not indicate any basis for the rejection of this hypothesis.

The second subsidiary problem had to do with the effect of promotion policy upon the general level of school achievement. Using as an index of rigidity or leniency of promotion policy the ratio of the number of pupil years of failure for the pupils enrolled in grade seven to the total number of pupils enrolled in that grade, two extreme groups of schools were identified, one a rigid group and the other a lenient group in terms of promotion policy. No school for which this index value fell below 0.2 was included in the rigid group. The mean value of this index for the twenty-five schools classified in this group was 0.3. The twenty-eight schools classified as lenient in promotion policy all had index values of zero. Using the mean score on the over-all composite for all students in each seventh grade class as a criterion of the level of achievement for a school, a simple *t*-test was applied to determine whether or not the criterion means differed significantly for these two types of schools. While the difference obtained favored slightly the schools having a lenient promotion policy, it was not statistically significant.

The third subsidiary problem had to do with the effect of promotion policy upon variability of pupil achievement. The two groups of schools used in connection with the foregoing problem were again employed. The standard deviation of the over-all composite scores for all pupils in each seventh grade class was used as a criterion of pupil variability within a given school. The means of these standard deviations were almost identical for the two types of schools.

The final subsidiary problem had to do with the effect of promo-

tion policy upon overageness. The two groups of schools used in the preceding two problems were again employed. The per cent of overage seventh grade pupils was used as a criterion of the extent of overageness prevalent in a given school. The means of these per cents for the rigid and lenient promotion policy schools were 17.3 and 6.4 per cent, respectively. The difference (10.9 per cent) between these means was statistically significant.

SUMMARY OF FINDINGS

The findings based on all comparisons as well as the subsidiary problems are summarized below. As is usually true of attempts to summarize complex results, there is considerable oversimplification. Special caution should be exercised with respect to generalizing from reported typical gains expressed in units of months. At least, it should be recognized that these units arise from a particular grade-equivalent scale associated with a particular test battery applied to a sample of children from a particular population.

(1) Failed pupils typically gain approximately only six months in educational progress during the repeat year and still fail to achieve the norm for the grade involved.

(2) Failed pupils typically gain approximately one year and three months in educational progress during the two years following failure and still fail to achieve the norm for the grade involved.

(3) During the year following failure, the educational progress of failed pupils is typically about four to six months less than that of matching promoted pupils.

(4) The educational progress of failed pupils during the two years following failure is not significantly greater (perhaps of the order of one to three months) than that made by promoted matchees during the single year spent in the next higher grade.

(5) The educational progress of seventh grade pupils who have experienced failure once is typically on a par with that of matched promoted seventh grade pupils who have spent one year less in school.

(6) The educational progress of seventh grade pupils who have experienced failure once is typically eight months less than that of matched promoted eighth grade pupils who have spent the same number of years in school.

(7) The educational progress of seventh grade pupils who have

experienced failure once is typically not affected by the grade in which the failure was experienced.

(8) The general level of achievement of a school's seventh grade class is not significantly affected by the rigidity or leniency of its promotion policy.

(9) The variability of pupil achievement in a school's seventh grade class is not affected by the rigidity or leniency of its promotion policy.

(10) The per cent of overage seventh grade pupils is significantly greater (perhaps of the order of ten per cent) in the case of schools having rigid promotion policies than in the case of schools having lenient promotion policies.

CONCLUDING REMARKS

Failure, in the form of non-promotion, as a device to ensure greater mastery of elementary school subject matter does appear justifiable in the light of the findings of this investigation. From the results reported, it would seem that slow learning children who are required to repeat a grade and slow learning children who are promoted, ultimately perform at about the same level when this performance is measured in the same higher grade, in spite of the fact that the failed pupils have each spent an added year in attaining this higher grade. It is not the intent to imply that a child should never be failed as he progresses through elementary school. However, if the consideration is solely a matter of educational achievement, it does seem clear that little is gained by requiring the repetition of a grade.

It should, of course, be clearly recognized that many other variables not studied in this investigation enter into the problem of non-promotion. Among these are personal and social adjustment or maturity, chronological age, and physical maturity. Also, the possible increase in cost of instruction accompanying a rigid promotion policy together with the needed adaptation of curricular offerings are factors which cannot be ignored in the determination of a policy. The problem is far from simple. It would seem that the critics referred to at the outset of this paper would, at least, be well advised to proceed with somewhat more of the caution and care that is characteristic of the type of scholarship they claim to represent.

REFERENCES

- 1) Garth H. Akridge, *Pupil Progress Policies and Practices*. Contributions to Education, No. 691, New York, N. Y., Teachers College, Columbia University, 1937.
- 2) R. D. Anfinson, "School Progress and Pupil Adjustment," *Elementary School Journal*, 41: 507-514, 1941.
- 3) Grace Arthur, "A Study of the Achievement of Sixty Grade I Repeaters as Compared with that of Non-Repeaters of the Same Mental Age," *Journal of Experimental Education*, 5: 203-205, 1936.
- 4) Bernard Iddings Bell, "Our Schools—Their Four Grievous Faults," *The Readers Digest*, January, 1951 (Condensed from *Life*, October 16, 1950).
- 5) Walter W. Cook, *Grouping and Promotion in the Elementary School*. Minneapolis, University of Minnesota Press, 1941.
- 6) Ann L. Crockett, "Lollipops vs. Learning," *Saturday Evening Post*, March 16, 1940.
- 7) Eugene S. Farley, "Regarding Repeaters, Sad Effects of Failures Upon the Child," *Nations Schools*, 18: 37-39, 1936.
- 8) John I. Goodlad, "Research and Theory Regarding Promotion and Non-promotion," *Elementary School Journal*, Vol. 53, November, 1952, 150-155.
- 9) *Iowa Tests of Basic Skills*. Boston, Houghton Mifflin Company.
- 10) Robert E. Larson, *Age-Grade Status of Iowa Elementary School Pupils*. Unpublished Doctoral Dissertation, State University of Iowa, 1955.
- 11) Henry J. Otto and Ernest O. Melby, "An Attempt to Evaluate the Threat of Failure as a Factor in Achievement," *Elementary School Journal*, 35: 588-596, 1935.
- 12) W. T. Wood, "How Well Are Our Schools Doing the Job?," *Vital Speeches of the Day*, March 1, 1952.

BOOK REVIEWS

ERNST W. SWANSON AND JOHN A. GRIFFEN, Editors. *Public Education in the South Today and Tomorrow*. Chapel Hill: The University of North Carolina Press, 1955, pp. 137.

In 1953 Harry S. Ashmore undertook to direct a study of biracial education in the United States. As a result there was published *The Negro and the Schools*. The present volume is an extension and refinement of the voluminous data gathered for that study. All this research and these publications were supported by the Fund for the Advancement of Education. Contribution to the original study and to this volume was made by a very considerable number of research workers and assistants. The result is "a handbook on statistics related to the South's educational problems."

The introductory chapter gives a brief summary of the findings. Chapters 2-5 consider the people of the South—population, growth and movement of the population; the children of the South—school-age population and attendance; some qualitative measures of southern schooling; and what the South is spending for schooling. Chapters 6-9 deal with the projection of school-age population, enrollment and daily attendance; the price of equality and improvement; the projection of southern income; and the South's ability to pay. Ten appendices explain techniques of the research, dealing with the selection of the sample, methods of projection of enrollment, of capital deficit, of income projection, etc.

The investigation assumes that, apart from integration, education of the South is faced with four major problems: equalization of current expenditures; modern housing; improvement of instruction; more classrooms and teachers. Comparisons are made between the South and the non-South, the South being defined for this study as the States of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia.

The authors clearly indicate that educational problems cannot be wholly determined in quantitative terms and point to qualitative differences that need to be considered. A dollar spent in one city or state cannot be considered as being of equal educational value as a dollar spent in some other city or state. However, careful statistical results permit insights that could not be had accurately without them.

The progress in education made in the South is heartening; the conclusion suggested in this study is that what needs to be done now and in the future is possible. No position is taken as regards integration; the measurements of deficits and projected expenditures have been figured in terms of the present dual system. It is thought that costs for the individual states would be "approximately the same, whether accomplished within a system of segregation, integration, or a combination of the two." But the people interested should see this most expensive function in its entirety and with as much accuracy as possible.

This book gives an over-all view with as much precision as may be expected in this complicated field of investigation. It is a handbook and should be a required book for those who are trying to solve the problems of schools and education of the South.

A. S. EDWARDS

The University of Georgia

CECIL V. MILLARD. *School and Child: A Case History*. East Lansing: Michigan State College Press, 1954, pp. 221. \$3.75.

In the author's own words this book has three objectives: (1) it "attempts to bring added light to the knowledge of development sequences of the preadolescent." (2) it "attempts to refine many of the broad principles and generalizations as to how the elementary school child grows and develops . . ." and (3) it illustrates "with one single case principle and generalization of the child's growth and development." These three objectives have been achieved with appreciable success.

Studies in human development have relatively neglected the school-age child from six to eleven—at least more attention has been devoted to the infant, preschool, and adolescent periods. For a number of years longitudinal studies have been made on school children by the Child Development Laboratory of Michigan State University under the author's direction. Data secured on seventy children who were available for the full period included measures of physical growth, mental development, achievement in school subjects, general development from the Courtis Motor and General Development Tests, and social status changes. In addition an annual case inventory was prepared from the formal data as well as from reports of student observers of children, home study, child's work samples, and interviews with the child.

The major portion of the present book is devoted to the study of Patricia as a continuing example of growth, about whom general principles are discussed. Six chapters of Part 1 describe the "mode" of each of the first six grades and detail the behavior of the illustrative case. Part 2 analyzes the changes in physical and mental development, academic achievement, personal-social development, and personal adjustment of Patricia. In the two chapters of Part 3, data presented earlier are used in an analysis of developmental patterns and the problems of studying child development.

For the reader, Patricia remains a real, normal girl who grows and in growing shows changing behavior in her school society. Tables of data and curves of growth help to condense material but the interrelations of all the variables of growth are seen most clearly in the picture drawn of the illustrative case for each grade, and then from the point of view of the several developmental areas. In over-all view the data lead to a picture of cyclic development with patterns showing changes at about seven years, ten years, and at the start of adolescence.

While essentially monographic in nature this book is written in a style that would make it an acceptable textbook or supplementary text in courses in child development. In particular it is valuable for the future teacher who can profit so greatly from an adequate understanding of the development of the average child.

C. M. LOUITT

Wayne University

Publications Received

- Actas Del Primer Congreso Argentino De Psicologia. Volumen 1, Organizacion 1, Problemas Historicos Y Epistemologicos De La Psicologia. Ministerio De Educacion De La Nacion: Universidad Nacional De Tucuman, 1955, pp. 380 (paper).*
- DONALD K. ADAMS. *The Anatomy of Personality*. Garden City, New York: Doubleday & Company, Inc., 1954, pp. 44 (paper). \$0.85.
- YRJO AHMAVAARA. *Transformation Analysis of Factorial Data and Other New Analytical Methods of Differential Psychology with Their Application to Thurstone's Basic Studies*. Helsinki, 1954, pp. 149 (paper).
- THERON ALEXANDER. *The Adult-Child Interaction Test: A Projective Test for Use In Research*. Champaign, Ill.: Child Development Publications, 1955, pp. 40 (paper).
- ROBERT M. ALLEN. *Psychological Tests and Measurements*. Coral Gables, Florida: University of Miami Press, 1954, pp. 115 (paper). \$1.25.
- FLOYD H. ALLPORT. *Theories of Perception and the Concept of Structure*. New York: John Wiley & Sons, Inc., 1955, pp. 709. \$8.00.
- GABRIEL A. ALMOND. *The Appeals of Communism*. Princeton, New Jersey: Princeton University Press, 1954, pp. 415. \$6.00.
- MILLIE ALMY. *Child Development*. New York: Henry Holt and Company, 1955, pp. 490. \$4.50.
- DEAN C. ANDREW AND LESTER N. DOWNING. *120 Readings in Guidance*. Lincoln: University of Nebraska Press, 1955, pp. 341 (paper).
- Applied Physics for High School*. Board of Education of the City of New York. Curriculum Bulletin, 1952-53 Series, Number 11, pp. 143 (paper).
- HARRY P. BAHRICK. *Teacher's Manual for "Psychology" by Wickens and Meyers*. New York: The Dryden Press, 1955, pp. 70 (paper).
- ALFRED L. BALDWIN. *Behavior and Development in Childhood*. New York: The Dryden Press, 1955, pp. 619. \$6.25.
- HAROLD BASOWITZ, HAROLD PERSKY, SHELDON J. KORCHIN AND ROY R. GRINKER. *Anxiety and Stress*. New York: McGraw-Hill Book Company, Inc., 1955, pp. 320. \$8.00.

- FRANZISKA BAUMGARTEN. *Die Regulierungs-Krafte im Seelen-Leben*. Bern: A. Francke Ag. Verlag, 1955, pp. 138 (paper).
- HOWARD BECKER AND REUBEN HILL, Editors. *Family, Marriage and Parenthood*. Second Edition. Boston: D. C. Heath & Co., 1955, pp. 849. \$6.25.
- LEOPOLD BELLAK. *The Thematic Apperception Test and The Children's Apperception Test in Clinical Use*. New York: Grune & Stratton, 1954, pp. 282. \$6.75.
- A. CORNELIUS BENJAMIN. *Operationism*. Springfield, Ill.: Charles C Thomas, 1955, pp. 154. \$4.00.
- IRWIN A. BERG. *Student's Workbook: For Use with "Psychology" by Wickens and Meyer*. New York: The Dryden Press, 1955, pp. 181 (paper). \$1.35.
- MORROE BERGER. *Racial Equality and the Law*. Paris: UNESCO, 1954, pp. 76 (paper).
- ARNOLD BERNSTEIN. *On The Nature of Psychotherapy*. Garden City, New York: Doubleday & Company, Inc., 1954, pp. 36 (paper). \$0.85.
- INEZ BERTAIL. *Complete Nursery Song Book*. New York: Crown Publishers, 1954, pp. 152. \$3.00.
- BRUNO BETTELHEIM. *Truants from Life*. Glencoe, Ill.: The Free Press, 1955, pp. 511. \$6.00.
- HOWARD BRAND. *The Study of Personality. A Book of Readings*. New York: John Wiley & Sons, Inc., 1954, pp. 581. \$6.00.
- LEON BRODY AND HERBERT J. STACK, Editors. *Highway Safety and Driver Education*. New York: Prentice-Hall, Inc., 1954, pp. 464. \$4.50.
- JAMES I. BROWN AND G. ROBERT CARLSEN. *Brown-Carlson Listening Comprehension Test*. Yonkers-on-Hudson, N. Y.: World Book Company (paper).
- CYRIL BURT. *The Subnormal Mind*. New York: Oxford University Press, 1955, pp. 391. \$5.00.
- ROBERT CALLIS, PAUL C. POLMANTIER AND EDWARD C. ROEBER. *A Casebook of Counseling*. New York: Appleton-Century-Crofts, 1955, pp. 352. \$5.00.
- STELLA S. CENTER AND WALTER N. DUROST. *Center-Durost Literature Acquaintance Test*. Yonkers-on-Hudson, N. Y.: World Book Company (paper).
- Child, Youth and Family Welfare, No. 1*. United Nations (Legisla-

- tive and Administrative Series, Volume III, 1951) New York: Columbia University Press (paper). \$1.75.
- ALBERT K. COHEN. *Delinquent Boys: The Culture of the Gang*. Glencoe, Illinois: The Free Press, 1955, pp. 202. \$3.50.
- ARTHUR P. COLADARCI, Editor. *Educational Psychology: A Book of Readings*. New York: The Dryden Press, 1955, pp. 656. \$3.90.
- ROBERT FLOYD CORDER. *A Factorial Approach to Anti-Democratic Attitudes*. Lafayette, Ind.: Purdue University, Division of Educational Reference, 1954, pp. 42.
- FAY L. COREY. *Values of Future Teachers*. New York: Bureau of Publications, Teachers College, Columbia University, 1955, pp. 146. \$3.50.
- LEE J. CRONBACH, Editor. *Text Materials in Modern Education*. Urbana, Ill.: University of Illinois Press, 1955, pp. 216 (paper). \$2.50.
- PAUL K. CROSSER. *The Nihilism of John Dewey*. New York: Philosophical Library, 1955, pp. 238. \$3.75.
- WILLIAM M. CRUICKSHANK. *Psychology of Exceptional Children and Youth*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1955, pp. 594. \$6.50.
- WENDELL W. CRUZE. *Psychology in Nursing*. New York: The Blakiston Division, McGraw-Hill Book Company, Inc., 1955, pp. 494. \$5.50.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

May, 1956

Number 5

SOME EFFECTS OF SHORT TERM TRAINING IN READING UNDER CONDITIONS OF CONTROLLED MOTIVATION

JAMES C. REED¹

University of Chicago

In recent years a number of attempts have been made to evaluate reading improvement training given to college students. By and large, the results of these studies have agreed in showing significant increases in rate. Some writers report a gain beyond chance in comprehension as measured by standardized tests. A few studies have indicated that such training has yielded measurable change in over-all academic proficiency as reflected by an improvement in grade point average. These results are encouraging, and at times provide information concerning the nature of the reading process. However, too often it is not at all clear that the reported gains were due to the interpolated remediation. Errors in methodological design and misuse of statistical techniques have imposed serious limitations on the generalizations that can be drawn from many of the studies as well as their validity. The two most serious weaknesses have been, first, a failure to equate the experimental and control groups on the basis of initial motivation, and secondly, a failure to randomly divide the participants. The difficulties encountered have been that volunteers have been paired with non-volunteers thereby permitting a bias. Also, the statistics t and F have been used to test the gains, and these values are based on statistical theory which assumes randomly chosen observations.

McGinnis' (8) experimental cases were volunteers to the reading clinic. Her control group was selected from the college at large, and though the two were carefully matched on the basis of sex, status in college, scholastic aptitude, and level of reading, they were not

¹ On leave of absence from Wayne University.

equated for initial motivation and the principle of random selection was violated. She found that the trainees exceeded the untrained in grade point average.

Kilby (6) also reported that instruction in reading resulted in significantly higher grades. He attempted to control for motivation by including among his matching criteria a "predicted grade score." It is questionable whether a predicted grade score derived from College Board Entrance Examinations, an achievement test, and adjusted high school grade-point is an adequate measure of motivation, and its applicability as such is suspect since Kilby's trained group were apparently volunteers and the controls picked from the general college population.

In Barbe's study (2), only those who elected reading were used. His experimental group consisted of the first twenty-five volunteers who were able to include the course in their schedules. The controls were the first twenty-five who could not so modify their program. While he reported significant gains in speed of reading and in honor point average for the experimental subjects, there is no evidence to indicate that the two groups were equivalent in intellectual ability. In addition, the method Barbe used for computing the significance of the rate gains will not yield a precise measure of the net shift in rate of reading. Furthermore, the possibility of differential motivation exists and the subjects were not randomly assigned. Robinson (11) likewise used only volunteers. However, the design was better in that control group which consisted of those who agreed to take the course at a later date. Unfortunately more than twenty-five per cent of them did not return for the class. Though Robinson's two groups were not based on random assignment, there is reason to believe that the factor of motivation was more nearly equated than in the studies heretofore mentioned. It is interesting to note that the grade point differences were not statistically significant although they approached being so. Mouly (10) saw the necessity of randomly sorting his subjects into experimental and control groups. He encountered difficulty in keeping motivation equal. Some of the students required to take the reading training refused to do so. Moreover, he found significant difference in grade point only when the experimental group was limited to those members who successfully completed the course. When the drop-outs and failures were included, the differences were no longer significant.

Such procedures as have been described suggest that the reported gains may be a function of characteristics of the participants rather than of treatments. The magnitude of gain seems inversely related to the degree of precision with which the experimental and control groups were matched. It may be that the traits which lead undergraduates to seek reading instruction are the same ones that result in higher grades than their less ambitious peers.

In the present experiment the hypothesis to be tested was that under conditions of controlled motivation, intensive training in reading and study skills would yield significant gains in rate, vocabulary, comprehension, and honor point average.

PROCEDURE

The subjects used were taken from the entering class of freshman hospital contract nursing students at Wayne University. From data contained in the entrance examination, the cases were matched by pairs on the basis of their ACE Q score (1) and their performance on the vocabulary and comprehension sections of the Diagnostic Survey Test, Form A (3). It was possible to find thirty-six individuals, eighteen pairs, whose intra-pair difference was no more than 5 raw score points on each of the three variables. The means for these thirty-six on ACE Q score, on comprehension, and on vocabulary were respectively at the twenty-fifth, fifty-first, and fiftieth percentile ranks on Wayne University freshman norms. By randomly assigning each member of each of the eighteen pairs into an experimental and control group, the differential motivation associated with the use of volunteer subjects was precluded. To determine whether there was comparability on the matching variables, means and standard deviations were computed. The difference between the means of the two groups on each of the three factors did not exceed 0.6 of a point. The sigmas were likewise identical.

After the experimental and control subjects had been selected, the entire class of freshman nurses was administered the Untimed Comprehension Test (3a) and the Vocabulary Test from the Diagnostic Battery (3b). On neither of these measures did the mean difference between the experimentals and controls exceed chance. It was announced that the course in reading improvement was required for student nurses, that there would be two sections—one

now and the other at a later date—and that their placement into a section would be purely by chance.

The reading class was composed of twenty-five members, eighteen of whom were the experimental subjects. The other seven were randomly chosen from those for whom no pair could be found at the time of the original matching. The training consisted of practice in rate of reading, drill on comprehension exercises, instruction in vocabulary, and discussion of study techniques. Glock's (5) *The Improvement of College Reading* was used as a text for the course supplemented by a wide variety of other material. The group met for an hour three times a week for nine weeks—a total of twenty-seven hours.

At the close of the training, alternate forms of the original tests were administered along with the Survey Section, Form B. A week later the control group and the remaining unpaired portion of the nursing class were separately given the same post-tests. The examiner stated that the second session of the course was to be offered during the spring semester and that at this time he wanted to determine how much gain was made in reading by students when

TABLE I.—DIFFERENCES IN GAIN (G) MADE BY EXPERIMENTAL AND CONTROL GROUPS

$N = 18$

Tests	Experimental					Control					Differences		
	Beginning		Ending			Beginning		Ending			$DG(e-c)$	σ diff.	t
	M	σ	M	σ	G_e	M	σ	M	σ	G_c			
Survey section													
Rate	268	33.25	353	59.83	85	249	39.63	285	42.69	36	49	20.20	2.44*
Vocab.	43	6.76	48	6.83	5	43	6.71	39	7.80	-4	9	2.46	3.61†
Comp.	30	4.39	30	2.85	0	30	5.30	31	3.42	1	-1	1.02	<1
Sec. 1	110	12.02	108	13.76	-2	116	7.99	109	9.76	-7	5	2.75	1.80
Vocab.													
Sec. 2	35	5.80	34	4.92	-1	37	5.20	35	4.78	-2	1	1.51	<1
Comp.													
Honor point			2.39	0.79				2.49	0.68		-0.1	0.19	<1

* Significant at two per cent.

† Significant beyond one per cent.

they did not receive special training, because if the gains were great enough, it was possible that the course would not be required.

The net gains in rate, vocabulary, and comprehension were calculated for each subject. The differences in net gain within each pair of subjects were computed for each of the three dependent variables. Using these distributions of differences, *t* tests were run. At the conclusion of the fall semester, the honor point averages of the subjects were obtained, and the differences were similarly evaluated. (9, p. 78).

Seven months after the end of the training period, the entire group was reassembled, and administered Form C of the Survey Section, and the same forms of the other tests that had been given to them initially. The statistical analysis of these delayed post-test results was done on the distribution of intra-pair difference scores of the experimental and control subjects.

RESULTS

As is contained in Table I, on the Survey Section of the Diagnostic Reading Tests, the experimental subjects made statistically significant gains over their paired controls in rate of reading, 0.02 per cent level, and on the vocabulary part, beyond the 0.01 per cent level. There were no significant differences found between the experimental and control groups on the vocabulary tests from Section 1 of the Diagnostic Battery, in comprehension, and in honor point average. In addition to the *t*, a nonparametric sign test as described by Dixon and Massey (4) was done for the pairs on each of the dependent measures. For rate of reading, the sign test was significant beyond the 0.01 confidence level. The remaining signs were not significant.

From Table II it can be seen that at the end of seven months the experimental group maintained its superiority in rate. The significant *t* between the two groups in the vocabulary section of the Survey Test was no longer present and none of the remaining differences exceeded chance. As before, only the sign test for rate of reading scores was significant.

CONCLUSIONS

The manner by which the subjects were selected for this experiment minimized the possibility that the experimental group initially had a higher degree of interest in the course or were more

TABLE II.—NET DIFFERENCES (D) BETWEEN EXPERIMENTAL AND CONTROL GROUPS SEVEN MONTHS AFTER TRAINING

 $N = 14$

Tests	Experimental		Control		Differences		
	M_e	σ	M_c	σ	$DM(e-c)$	σ diff.	t
Survey section							
Rate	320	33.25	283	22.66	37	9.05	4.07*
Vocab.	40	4.34	40	4.47	0	1.61	<1
Comp.	30	4.53	31	3.05	-1	1.46	<1
Sec. 1 Vocab.	116	9.73	119	7.46	3	2.48	1.10
Sec. 2 Comp.	35	5.97	37	5.84	-1.5	2.20	<1

* Significant beyond one per cent.

strongly motivated than the control group as frequently happens when volunteers are employed. Judging from the response of the nurses in the classes, their number of absences, the frequency of late comers, their postural indications of attention, it is believed that the group interest and motivation was equal to that of an average introductory "required" course. The nurses were somewhat inclined to be phlegmatic and they did what they were told without too much question. Consequently, it is felt that the changes in reading skills are due to the training and not to antecedent conditions peculiar to the experimental subjects.

The conclusions and generalizations that may be drawn from these results are necessarily limited due to the small size of the N . It should be remembered that the sample is not representative of a college population. However, it would appear that the following conclusions are justified.

(1) The significant increase in rate of reading without change in comprehension or vocabulary suggests that rate is an independent factor in the reading process. Furthermore, rate will increase as a result of training and this increase has some degree of permanence.

(2) Since a significant difference was found on the timed vocabulary test of the Survey Section but not on the more liberally timed ones of Section 1, it appears that the vocabulary section of the Diagnostic Survey Test is in part a measure of rate. Such an interpretation is further supported by the fact that the difference on the Survey vocabulary test between the experimental and controls was not significant on the delayed post-tests.

(3) Under appropriate matching and equating of initial motivation, short-term training in reading does not yield material differences in fundamental comprehension and vocabulary skills.

(4) It would likewise appear that changes in honor point averages are not brought about by reading training. Grades are a function of a complex interaction of many variables, and are in part due to habits and abilities of long standing. It would be rather surprising if as the result of a nine-weeks course, there were substantial differences in academic achievement.

(5) It may be that failure to achieve significant changes in grade point is due to the specialized curriculum in which the nurses were enrolled. Their courses were mainly in the biological and physical sciences where quite intensive and comprehensive reading is required. It is possible that the training as described herein would be beneficial in terms of an increase in honor point for students enrolled in the social sciences and humanities. Credence is given such an interpretation by the works of Sheldon (12), Kilby (6), and Kingston (7). However, the writer does not share the view that grade points are a sole or necessarily even a useful criterion in evaluating a reading program.

(6) It is recognized that discrepancies between the results of this and other experiments may possibly be due to differences in instructional procedures, the length of time the classes were offered, the type of students, and even the competence of the instructors. Yet, it is felt that this experiment has demonstrated a need for more careful evaluations of reading programs at the college level. More attention should be directed to the appropriate matching of experimental and control subjects. The precision of the experiments can be increased by the use of appropriate estimates of error. Finally, repetition of such studies must be done in order to randomize the types of error not possible to eliminate or randomize by any single experiment.

REFERENCES

- 1) American Council on Education, *Psychological Examination*. Cooperative Test Division, Educational Testing Service, Princeton, New Jersey.
- 2) Walter H. Barbe, "The Effectiveness of Work in Remedial Reading at the College Level," *Journal of Educational Psychology*, 43: 229-237, 1952.
- 3) *Diagnostic Reading Tests*. Survey Section. Committee on Diagnostic Reading Tests, Kingscott Apartment, 3 G, 419 West 119th Street, New York 27, New York.

- 3a) *Diagnostic Reading Tests*, Section II, Comprehension, Part I Silent.
- 3b) *Diagnostic Reading Tests*, Section I, Vocabulary.
- 4) Wilfred J. Dixon and Frank J. Massey, Jr., *Introduction to Statistical Analysis*, pp. 247-254. New York, McGraw-Hill, 1951.
- 5) Marvin D. Glock, *The Improvement of College Reading*. Chicago, Houghton-Mifflin Co., 1954.
- 6) R. W. Kilby, "The Relation of a Remedial Reading Program to Scholastic Success in College," *Journal of Educational Psychology* 36: 513-534, 1945.
- 7) Albert J. Kingston and George E. Clay, "The Effectiveness of Reading Training at the College Level," *Journal of Educational Research* 48: 467-471, 1955.
- 8) Dorothy McGinnis, "Corrective Reading: A Means of Increasing Scholastic Attainment at the College Level," *Journal of Educational Psychology*, 42: 166-173, 1951.
- 9) Quinn McNemar, *Psychological Statistics*, p. 78. New York, J. Wiley, 1949.
- 10) George J. Mouly, "A Study of the Effects of a Remedial Reading Program on Academic Grades at the College Level," *Journal of Educational Psychology*, 43: 459-466, 1952.
- 11) H. A. Robinson, "A Note on the Evaluation of College Remedial Reading Courses," *Journal of Educational Psychology*, 41: 83-96, 1950.
- 12) William Sheldon, "An Evaluation of an Experimental Reading Program for Medical Students," *Journal of Educational Psychology* 39: 298-303, 1948.

A NOTE ON FACTOR MODELS

CHESTER W. HARRIS

University of Wisconsin

It is important to distinguish, as Maxwell (9) has done, between the choice of a factor model and the choice of computation procedure. His preference, like that of many others (2, 3, 4, 7, 11), is for the communality principle. The purpose of this note is to emphasize two points: (1) that two well-known systems of factor analysis—i.e., factor models—though distinguishable, are related in a regular fashion, and (2) that computation procedures which are readily performed on digital computers, such as those described by Wrigley and Neuhaus (12), may be utilized to secure a solution utilizing the communality principle.

FACTOR MODELS

The distinction between two systems of factor analysis, or factor models, can be visualized by considering the trivial problem of only two variables. One approach is that of factoring the two-by-two matrix consisting of units in the principal diagonal and the correlation between the two variables in the off-diagonal positions. This takes as a complete initial orthogonal solution a first factor midway between the two test vectors but lying in the plane of these two vectors, and a second factor at right angles to this first factor but again lying in the plane of these two test vectors. This approach locates the factors *within the test space*. A second approach is given by the communality principle, which corresponds to factoring correlations with communalities in the principal diagonal. Three factors are identified, two of which are unique to the two variables. The single common factor is located at the intersection of two orthogonal planes, each of which passes through one of the test vectors, and consequently is located *outside the test space*. It also is true that the unique factors are located outside the test space. These two approaches, then, differ in that for one the factor space is defined as identical with the test space, whereas for the other the common-factor space is defined as distinct from the test space.

A number of qualifying comments should be introduced at this

point. For example, it should be remarked that the two approaches give identical results under the condition of perfect correlation between the two variables, and the effects of using covariances rather than correlations should be considered.

It has been shown (6) that this distinction in factor models corresponds to the distinction between factoring the given data, such as a normalized set of test scores for N individuals, and factoring a projection of this matrix of data. Both procedures imply a separation of the given data into two parts, which is the basic statistical operation in analysis of variance and in linear regression. The communality principle in factor analysis can therefore be formulated as a straight-forward statistical problem. Anderson and Rubin (1), Lawley (8), and Rao (10) have discussed alternative ways of separating the given data into two parts and the types of solutions these imply.

Given the distinction between a set of orthogonal factors located within the test space and one or more common factors located outside the test space by using the communality principle, the interesting question of the relationship between these two distinct sets of factors arises. As an illustration, consider a two-variable matrix:

$$\begin{vmatrix} 1.00 & 0.64 \\ 0.64 & 1.00 \end{vmatrix}$$

The communality solution yields a single common factor, here labeled G , and the complete principal-axis (or centroid in this case) solution yields two common factors, here labeled I and II:

Variable	G	I	II
1	0.8	$\sqrt{0.82}$	$\sqrt{0.18}$
2	0.8	$\sqrt{0.82}$	$-\sqrt{0.18}$

The transformation that projects the complete set of factors, I and II, into the single common factor, G , can readily be computed for this simple problem. This transformation is simply a single row vector with a non-zero entry, consisting of $\sqrt{0.64}$ divided by $\sqrt{0.82}$, in the first cell and a zero in the second cell. More generally, it can be shown (5) that for population values of the correlations for which a communality solution exists, such a trans-

formation always exists and consists only of real numbers. When this relationship is considered in connection with the problem of estimating factors, it turns out that the transformation can be defined algebraically, and that common factors solved for by this transformation satisfy Lawley's maximum likelihood requirement and are a first approximation to Rao's canonical factors.

This relationship between two systems of factor analysis therefore, in one sense, blurs the distinction in factor models emphasized by Maxwell, since it shows the mode of transforming one to the other; yet at the same time it helps sharpen the distinction by specifying the projective nature of this transformation.

COMPUTATION PROCEDURES

Implicit in this discussion is the notion that one might factor completely any given correlation matrix with units in the diagonal cells, and then by further calculation determine common factors as a transformation of this complete set of factors. (Variances and covariances might be used instead, and the resulting factors would be simply related to those determined by operating on the correlations.) This might be done by machine, especially since all of the factoring required by this process should be done by the principal-axis transformation, which is well-adapted to digital computing. However, this process actually is uneconomic compared with the direct attack on the estimation of common factors given by Rao (10) in his notion of canonical factors. His computation procedure also requires the principal-axis transformation of a matrix, and thus is well-adapted to digital computing.

Perhaps the best advice for the factor analyst today is, "Study Rao's canonical factor analysis procedure and whenever the matrix to be factored is of any size, arrange for machine computation, employing Rao's method." Operations appropriate to high-speed computing machines are employed, a "respectable" mathematical model is used, and appropriate tests of statistical significance are made available. In addition—and this is not insignificant—the communality principle is preserved.

REFERENCES

- 1) T. W. Anderson and Herman Rubin, *Statistical Inference in Factor Analysis*. Technical Report No. 30, Contract N6onr-251, Office of Naval Research. Stanford, California, Department of Statistics, Stanford University 1955.

2) Raymond B. Cattell, *Factor Analysis*. New York, Harper and Brothers, 1952.

3) H. J. Eysenck, "The Logical Basis of Factor Analysis," *The American Psychologist*, pp. 105-114, March, 1953.

4) J. P. Guilford, *Psychometric Methods*. New York, McGraw-Hill, 1954.

5) Chester W. Harris, "Relationships Between Two Systems of Factor Analysis," *Psychometrika*, (in press).

6) Chester W. Harris, "Separation of Data as a Principle in Factor Analysis," *Psychometrika*, pp. 23-28, March, 1955.

7) Karl J. Holzinger and Harry S. Harman, *Factor Analysis*. Chicago, University of Chicago Press, 1941.

8) D. N. Lawley, "The Estimation of Factor Loadings by the Method of Maximum Likelihood," *Proceedings, Royal Society of Edinburgh* pp. 64-82, 1940.

9) A. E. Maxwell, "Factor Models," *Journal of Educational Psychology*, p. 9 March, 1956.

10) C. R. Rao, "Estimation and Tests of Significance in Factor Analysis," *Psychometrika*, June, 1955, 93-111.

11) L. L. Thurstone, *Multiple-Factor Analysis*. Chicago, University of Chicago Press, 1947.

12) Charles Wrigley and J. O. Neuhaus, "Use of an Electronic Computer in Principal Axes Factor Analysis," *Journal of Educational Psychology*, pp. 31-41, January, 1955.

THE DAVIS-EELLS TEST AND READING SUCCESS IN FIRST GRADE

IVAN L. RUSSELL

Southern Illinois University

The Davis-Eells Test is a unique approach to the measurement of intelligence. Attention is focused upon the child's capacity for solving problems similar to true-life experiences. An attempt is made to minimize the influence of socio-economic differences. The purpose of the Davis-Eells Test is to "afford the teacher a measure of a pupil's mental capacity, of his basic resources of reasoning, insight, problem organization and thus to throw new light for her upon the pupil's true capacity." (1)

A study was conducted to test the hypothesis that the Davis-Eells Test provides important information regarding reading readiness. The plan of study included administration of the Davis-Eells Test to a randomly selected group of fifty first-grade pupils during the first week of school. The test was administered to groups of ten pupils each, and the directions were followed carefully. For comparative purposes the Stanford-Binet Test, Form L, was administered to each of the fifty pupils during the first two weeks of school. During the last two weeks of school in May the Gates Sentence and Paragraph Reading Tests were given. These sets of test scores were compared statistically with the Gates reading scores, which were obtained at the end of the first grade, as the criterion of reading success.

PRESENTATION OF DATA

Table I compares the means and standard deviations of IQ's earned on the Stanford-Binet and Davis-Eells Tests.

These differences between the means on the two tests are significant at the one per cent level. The difference between mean Davis-Eells scores for boys and for girls is significant at the five per cent level.

A Pearson product-moment correlation coefficient of $+ 0.48$ was obtained between the Davis-Eells IPSA (Index of Problem Solving Ability) and Stanford-Binet IQ scores. Stanford-Binet mental ages correlated with Gates reading ages obtained at the

TABLE I.—MEANS AND STANDARD DEVIATIONS FOR DAVIS-EELLS AND STANFORD-BINET SCORES

Group	N	Davis-Eells (IPSA)		Stanford-Binet (IQ)	
		Mean	S.D.	Mean	S.D.
Girls	28	95.3	12.2	110.0	14.5
Boys	22	87.9	12.4	104.5	14.6
Total	50	91.9	13.1	108.9	14.8

end of the school year are as follows: for boys, + 0.48; for girls, + 0.63; for total group, + 0.57.

The Davis-Eells test does not provide M.A. scores. These were derived by multiplying each pupils IPSA score, roughly equivalent to IQ, by his chronological age. The M.A.'s thus derived were correlated with end-of-the-year reading scores, just as in the case of the Stanford-Binet Test. The coefficients of correlation which resulted were: girls $r = 0.28$; boys $r = 0.08$; total group $r = 0.21$.

CONCLUSIONS

(1) IPSA scores obtained in this study involving first-grade pupils proved to be significantly lower than Stanford-Binet IQ's ($p = 0.01$).

(2) For this limited sample the Stanford-Binet test gives a better prediction of reading progress during the first year's instruction than the Davis-Eells Test.

REFERENCES

- 1) Allison Davis and Kenneth Eells, *Manual: Davis-Eells Test of General Intelligence or Problem-Solving Ability*, p. 2. New York, World Book Co., 1953.

THE MINNESOTA TEACHER ATTITUDE INVENTORY IN EVALUATING THE TEACHING OF EDUCATIONAL PSYCHOLOGY

MORRIS E. ESON¹

New York State College for Teachers, Albany, N. Y.

One of the important objectives of the professional courses in teacher education is the development and change of attitudes toward the learner and the teaching task. The teacher of such courses may be reluctant to measure the extent to which this objective is achieved because there have been no well-standardized instruments for the measurement of teacher attitudes and also because the connection between the content of the professional courses and any set of attitudes is not always clear. The study reported here represents an attempt to measure the attitude change which took place in one such course, the first in the professional sequence at New York State College for Teachers, Albany, N. Y. This course, Education 20, is generally taken by the student during the first semester of the sophomore year and follows a three-semester-hour course in General Psychology. It is a course in Educational Psychology and is described in the 1954-55 college catalogue as follows:

The study of basic concepts of human behavior as related to the educational situation. Special emphasis will be given to physical, social, emotional and mental development in early and later childhood; the study of the interaction of heredity and environment in the development of personality; the learning process including a consideration of reasoning, remembering, transfer of training and the various theories of learning. One two hour period each week will be spent in laboratory session to develop insight into human behavior by observation, preparation of reports and other laboratory devices.

The text used in the course was James L. Mursell's *Psychology for Modern Education*, W. W. Norton & Co., 1952. The writer met the classes twice each week for sixteen weeks. These class sessions

¹ The writer gratefully acknowledges the help of Owen Smith and Nancy D. Westover, graduate students at New York State College for Teachers, in the statistical analysis of the data.

were spent in lecture, discussion and demonstration. Each student observed five times in elementary school classrooms and five times in secondary school classrooms and was required to submit observation reports on seven of these observations. These reports were returned with comments by the instructor. The instructor's philosophy may be described as consistent with that which would be scored high on the Minnesota Teacher Attitude Inventory. The instructor's score on the Inventory was such as to place him at the ninety-fifth percentile on the norms established for secondary school teachers with five years of training.

The population used in the study consisted of seventy-five first-term sophomore men and women who were taking their first professional course in education, Educational Psychology. The data used in the study were collected as follows: a) During the first class session the Inventory was administered. b) After eight weeks an instructor-constructed objective test covering the material in approximately the first half of the text was administered. c) The Inventory was again given during the last class session. d) A final objective test, instructor constructed, was given in the final examination week. At each of the administrations of the Inventory, the students were assured that their scores would in no way influence their final grade. The response sheets on the Inventory were machine scored.

The rationale of the Inventory may best be described by a series of quotations from the Manual (1).

It is assumed that a teacher ranking at the high end of the Scale should be able to maintain a state of harmonious relations with his pupils characterized by mutual affection and sympathetic understanding.

The teacher should like the children and enjoy teaching.

The teacher and pupils should work together—with a feeling of security growing from a permissive atmosphere of freedom to think, act and speak one's mind with mutual respect for the feelings, rights and abilities of others.

At the other extreme of the Scale is the teacher who attempts to dominate the classroom.

Creating an atmosphere of tension, fear and submission.

There is a feeling of mutual distrust and hostility.

The teacher tends to think in terms of . . . the subject matter to be covered rather than in terms of what the pupil needs, feels, knows and can do.

He frequently has a submissive uncritical attitude toward authorities over him and a dominating, overbearing attitude toward subordinates.

The authors of the Inventory state that the items in the Inventory "discriminate sharply between teachers who have and those who do not have good rapport with pupils; examination of these items indicates that inferior teachers are essentially insecure socially." The factors suggested causing social insecurity are general appearance, failure in heterosexual adjustment, low social status of family, failure to be accepted socially in high school and the like. In teaching the course here being evaluated, no attempt was made to change these factors; obviously some of these could not be changed. In fact it would seem that the attitudes measured by the Inventory are resistant to change and that "the mere inculcation of better attitudes by instruction may not produce any change in behavior." It is to be speculated as to whether instruction in the principles of child development and the learning process may produce changes in attitudes as measured by the Inventory and, if so, whether these attitude changes would reflect also a change in behavior.

RESULTS

The mean score on the Inventory increased by 11.5 points, as seen in Table I, between the first and the second testing. The product moment correlation between the two sets of scores of the Inventory was found to be 0.69. The significance of the difference for correlated means was tested by obtaining the Standard Error which was found to be 2.58 giving a t value of 4.4, significant beyond the 0.01 level.

A correlation of +0.29 was found between the scores on the final content test and the second attitude inventory. This low

TABLE I—COMPARISON OF SCORES ON THE MINNESOTA TEACHER ATTITUDE INVENTORY TAKEN SIXTEEN WEEKS APART

	N	M	r	SD	Difference between Means	SE diff	t
MTAI first administration	75	41.4	0.69	26	11.5	2.58	4.4
MTAI second administration	75	52.9		28			

correlation between the course content and the attitudes measured on the inventory is further confirmed when the twenty-six students with the highest inventory percentile ranks are compared with the twenty-six lowest inventory scores; using letter grades obtained by these students in the course a contingency coefficient of $+0.34$ was obtained.

The mean score of the population when compared with beginning education juniors, secondary academic was found to be at about the fortieth percentile on the first testing and about the fifty-fifth percentile on the second testing of the Inventory.

Seventy-two per cent of the students in the study favorably increased their scores by more than five points whereas twenty per cent decreased their scores by more than five points.

DISCUSSION

Since the difference between the means on the Inventory was found to be statistically significant, one might be tempted to conclude that the course contributed to the improvement of attitudes. However, one might also conclude that the students had simply become test wise. The authors of the Inventory report two administrations of the Inventory with seven to ten days between testings and with standard instructions each time. The difference between means in that instance is found to be significant beyond the one per cent level of confidence. Our situation here parallels even more closely the sequence described by the authors in the Manual in which the subjects were encouraged to "fake good" on the second administration which followed the first by a four- to six-week interval. The authors report a t value of 4.13 which corresponds very closely to the 4.4 reported above.

The fact that correlation between the results on course content examinations and the Inventory scores is found to be low is further evidence that the difference between the Inventory mean scores, although statistically significant, is in reality not significant.

SUMMARY AND CONCLUSIONS

The Minnesota Teacher Attitude Inventory was administered to a group of seventy-five college sophomores at the beginning and at the conclusion of a course in Educational Psychology. The mean score was raised 11.5 points which is shown to be a statistically significant difference. The low correlation between scores on sub-

ject matter mastery and the attitude inventory as well as the possibility of faking on the attitude inventory cause the author to question the applicability of the Minnesota Teacher Attitude Inventory to the evaluation of courses which attempt to create and develop certain kinds of teacher attitudes.

REFERENCE

- 1) W. W. Cook, C. H. Leeds and R. Callis, *Minnesota Teacher Attitude Inventory Manual*. The Psychological Corporation, New York, N. Y.

REGENCY AND TYPE OF UNDERGRADUATE TRAINING AND DECLINE OF G.R.E. PERFORMANCE WITH AGE¹

R. TRAVIS OSBORNE and WILMA B. SANDERS

INTRODUCTION

Previous papers (1, 2, 3) by the writers have shown that there is a significant decline in mean Graduate Record Examination scores with increased age. In a study of one thousand, eight hundred and seven graduate students, the rate of loss varied both with age and subject-matter area. Over the nineteen to sixty-five age range the science profile age-score curves showed losses ranging from zero to seventeen per cent for men and from seven to twenty-five per cent for women. Achievement scores in literature held up well for both sexes over the entire age range. The question was raised "Are there factors operating other than age or in combination with age to produce these changes?" For example, regency and type of undergraduate training might affect the rate of profile score change with age. Students who earned baccalaureate degrees in the sciences not only might be superior initially in their major fields but also might show a less marked rate of achievement loss with age than non-science majors. It is also hypothesized that, regardless of age, recent graduates will test higher than members of earlier classes. The purpose of this paper, which is subject to the limitations inherent in a cross sectional study, is to present research evidence relevant to these points.

POPULATION

The one thousand, eight hundred and seven graduate students who were in attendance at the University of Georgia between February, 1946, and February, 1952, were subjects for the study. This group included the entire graduate student body except several who were not seeking advanced degrees. The seven professional schools and the five divisions of the College of Arts and Sciences were represented. Men outnumbered women one thousand one hundred and seventy-three to six hundred and thirty-four. The

¹ Based upon a paper presented at the meeting of the American Psychological Association, New York, September, 1954.

age range of subjects was from nineteen to sixty-five with a mean of 31.76 years.

RESULTS

In Table I, Graduate Record Examination profile scores are shown fractionated by time interval between the student's baccalaureate degree and his G.R.E., disregarding age at the time of the test. Thus it is possible in Table I to compare achievement profiles of recent graduates with those of students who received their A.B.'s four, eight, or up to twenty years previously. The losses in scores found when the G.R.E. results were fractionated by age (3) are attenuated when the data are broken down by recency of baccalaureate degree. Only on three profiles are significant differences noted in favor of the most recent graduates. Biology, chemistry, and physics scores are elevated for students who had received their degrees within three years of the time of testing. Four years after graduation, all inter-group differences on the chemistry and physics profiles as well as most differences in biology are insignificant (Table I).

Recency of undergraduate training is not a significant factor at any age level in the areas of mathematics, social science, fine arts, and literature. In fact on the literature profile the "old grads," class of 1923 and before, had significantly superior test results compared to the four other age groups (Table I).

To test the hypothesis that type of undergraduate training affects the rate of profile score change with age, Table II was prepared. It shows the means, standard deviations, and the number of science and non-science majors at each age level. B.S. students were found to be superior on the chemistry, physics, and biology profiles at practically every age level whereas non-science students excelled in social science and the humanities at all age levels. Declines appear to be sharpest for science majors on the science profiles. Science achievement loss for non-science majors is gradual with the means for the two curriculum groups becoming more nearly equal after age thirty-nine.

In order to establish an over-all test of significance these data were further subjected to analysis of variance with proper adjustment being made for disproportionality (4). The hypotheses to be tested by this design are: (1) The effects of student age on the profile scores of the G.R.E. are insignificant. (2) Variations in

TABLE II—TYPE OF BACCALAUREATE DEGREE IN RELATION TO GRADUATE RECORD EXAMINATION PERFORMANCE

Age group	Physics						Biological Sciences					
	Baccalaureate degree in science			Baccalaureate degree non-science			Baccalaureate degree in science			Baccalaureate degree non-science		
	No.	Mean St. Sc.*	S.D.	No.	Mean St. Sc.*	S.D.	No.	Mean St. Sc.*	S.D.	No.	Mean St. Sc.*	S.D.
19-22	46	51.09	10.97	82	49.67	8.87	46	58.33	9.66	82	52.41	8.28
23-26	196	54.15	9.73	211	49.62	8.01	196	56.08	9.25	211	50.67	8.33
27-30	188	53.33	10.70	159	48.83	8.98	188	54.07	9.66	159	49.75	7.99
31-34	91	51.93	9.62	112	49.28	8.87	91	51.17	9.43	112	47.96	9.60
35-38	67	49.99	10.32	131	49.40	9.73	67	46.53	8.97	130	47.11	10.11
39-42	50	50.17	8.76	97	44.62	9.41	50	47.80	8.91	97	44.27	8.45
43-46	26	45.22	8.60	62	47.79	11.72	26	44.41	10.23	62	46.61	10.92
47-50	12	50.39	9.57	39	45.84	8.66	12	47.40	8.79	39	42.36	8.16
51- up	13	38.81	5.91	38	42.89	9.62	13	37.78	8.56	38	41.42	9.66
Chemistry												
19-22	44	55.55	10.99	80	48.81	10.06	33	47.23	9.95	56	50.32	8.82
23-26	192	56.22	9.32	202	47.89	9.50	156	52.12	9.84	179	50.02	10.22
27-30	176	53.50	10.12	143	47.48	8.63	153	52.00	9.68	129	49.23	10.16
31-34	83	52.50	9.63	106	48.02	8.94	63	48.74	10.97	94	49.50	10.48
35-38	57	51.28	8.20	111	46.17	9.19	49	50.59	10.59	100	51.29	11.77
39-42	49	51.35	8.39	85	45.17	7.95	37	48.49	10.91	77	47.11	9.03
43-46	24	49.02	9.63	54	46.55	9.69	16	42.41	7.58	48	49.01	11.83
47-50	11	56.68	8.32	39	46.51	9.19	8	50.13	8.28	31	46.65	10.38
51- up	12	42.28	4.04	37	44.72	8.14	9	43.64	11.61	27	49.41	10.48
Mathematics												
Social Sciences												
19-22	46	47.65	8.81	82	52.16	10.00	46	49.01	8.30	82	55.02	9.95
23-26	196	47.08	7.98	211	52.93	9.79	196	46.63	8.52	211	54.16	9.67
27-30	188	47.09	8.96	159	51.68	10.88	186	45.45	8.90	159	52.59	10.60
31-34	91	47.46	9.69	112	51.97	10.31	91	45.97	9.07	112	51.62	10.99
35-38	67	48.09	10.05	129	52.58	11.35	67	45.78	8.96	130	51.74	9.84
39-42	50	47.49	8.91	97	50.68	10.05	50	47.32	8.46	97	51.90	9.07
43-46	25	49.57	9.02	62	51.74	10.78	26	47.26	7.86	62	54.02	11.04
47-50	12	46.54	7.46	39	47.97	8.29	12	44.87	7.03	39	48.88	8.02
51- up	13	46.78	6.94	38	50.79	9.38	13	47.04	8.90	38	51.27	10.00
Literature												
Fine Arts												
19-22	46	49.66	10.90	82	54.40	10.00						
23-26	195	47.55	8.39	209	54.23	10.20						
27-30	178	47.20	8.74	158	51.62	10.90						
31-34	88	46.04	9.05	110	51.79	10.75						
35-38	66	45.81	8.34	129	50.85	10.10						
39-42	48	46.91	8.24	96	49.27	9.60						
43-46	26	49.00	10.00	60	52.94	9.50						
47-50	12	42.72	8.74	38	45.31	7.49						
51- up	13	49.97	10.30	37	50.56	9.25						

* University of Georgia norms ($N = 1716$).

G.R.E. performance are not significantly related to the type of the students' undergraduate training. (3) Variations in rate of achievement loss with age are not significantly different for students with dissimilar undergraduate training. Table III gives the results of repeating this analysis for each of the seven profile tests.

On the basis of evidence presented in Table III the first hypothesis must be rejected for six of the seven profiles. Only in the case of social science is age not a significant determiner of G.R.E. scores. For the other six subject-matter areas, age increments are accompanied by a significant decrease in mean test scores.

Hypothesis two is untenable in six of seven instances. Curriculum differences are significant for all profiles except mathematics. Students with different types of undergraduate training differ in their retention of subject matter content evaluated by the G.R.E. profiles. Initially B.S. students test significantly higher in chemistry, physics, and biology than non-science majors; however, these differences are attenuated beyond age thirty-nine. Younger students with undergraduate A.B.s were superior on the social science, fine arts, and literature profiles but these differences, too, were less marked in the later age groups.

Age-curriculum interaction terms for the various profiles deserve special comment. Curriculum-age interaction values were not significant for social science, literature, and fine arts. This in part confirms hypothesis number three. For chemistry, physics, and biology, however, interaction is significant; that is, the superiority or inferiority of the two curriculums is more pronounced at certain age levels than at others. The rate of science achievement loss with age is not the same for both science and non-science undergraduate majors.

DISCUSSION AND SUMMARY

In previous studies which reported significant G.R.E. achievement losses with age it was suggested that recency of undergraduate training might be a factor determining these marked changes in achievement with increasing age. There is little evidence in the present study to support this argument. Four years after graduation all significant chemistry and physics achievement differences have been erased; most biological science differences have

also disappeared. On the non-science profiles, recency of undergraduate training is not significant at any age level.

The type of the student's undergraduate training is reflected in his achievement scores at all age levels and in the expected directions. Students who majored in the social sciences and humanities were superior on profiles representing their fields while science majors excelled on the chemistry, physics, and biology tests. Sharp, positively accelerated losses in mean scores for science majors on the science profiles almost wipe out their initial superiority by age thirty-nine.

The full influence of curriculum differences on achievement variations is not clear from the data reported in the previous section. It may be said, however, that the rate of achievement loss with age is differential for students with different types of training but after age thirty-nine most curriculum effects are insignificant.

IMPLICATIONS

Implicit in the experimental design of the Graduate Record Examination were two hypotheses: first, that graduate level subject-matter achievement tests could be used efficiently by deans and admission officers for the selection and evaluation of graduate students; two, that these tests would reveal to the students their strengths and weaknesses in broad areas of learning covered by the profiles. The original G.R.E. standardization procedure further implied that all graduate students regardless of age or type of undergraduate training could be evaluated in terms of the same "standard coin."

Selection and Evaluation. Admission officers using the G.R.E. for selection purposes should be cognizant of the differential decline of profile scores with age and type of previous training. Scores become much more meaningful when comparisons are made between individuals from the same chronological age group. In some areas there are influences that operate to favor the younger student, some areas are apparently unaffected by age changes, while some influences favor the older age group.

Counseling and Guidance. The individual student who uses G.R.E. profiles for evaluating his past learning experiences and for planning his educational future should not only be aware of

his relative standing among all graduate students but also he should know how he compares with age peers who have had the same type of undergraduate training. Recency of undergraduate training can not be entirely ignored in educational planning at the graduate level but that is not as significant a determiner of G.R.E. profile scores as age or type of undergraduate training.

REFERENCES

- 1) R. Travis Osborne and Wilma B. Sanders, "Comparative Decline of Graduate Record Examination Scores and Intelligence With Age," *The Journal of Educational Psychology*, Vol. 45, No. 6: October, 1954, 353-358.
- 2) R. Travis Osborne and Wilma B. Sanders, "Differential Decline in Graduate Record Examination Scores With Age," *Journal of Genetic Psychology* (in press).
- 3) R. Travis Osborne and Wilma B. Sanders, "Variations in Graduate Record Examination Performance By Age and Sex," *Journal of Gerontology*, Vol. 9, No. 2: April, 1954, 179-185.
- 4) James E. Wert, Charles O. Neidt and J. Stanley Ahmann, *Statistical Methods in Educational and Psychological Research*. New York, Appleton-Century-Crofts, Inc., 1954, 435 pp.

THE DEVELOPMENT OF NEW SCALES FOR THE APTITUDE AND ADVANCED TESTS OF THE GRADUATE RECORD EXAMINATIONS¹

MARGARET K. SCHULTZ and WILLIAM H. ANGOFF

Educational Testing Service, Princeton, N. J.

The Graduate Record Examinations offer, among other tests, an Aptitude Test, yielding a Verbal score and a Quantitative score, and a series of Advanced Tests in the usual major areas of college study—biology, chemistry, history, psychology and so forth. These tests are taken primarily by college seniors and graduate students in testing programs sponsored by their colleges, and by applicants to graduate schools which require or recommend the tests for admission.

The scaled score system for these tests and for the other tests in the Graduate Record Examinations program is one which defines a scaled score of 500 as the mean score for the particular standardization or reference group on which the scale is based, and 100 as the standard deviation of these scores. Such a score system has little inherent normative value, however, if the standardization group whose performance is the basis for the scale is not reasonably representative of groups who currently use the tests. In the case of the Aptitude Test, groups now being tested are far more heterogeneous and generally lower in ability than the original standardization group, which was composed of first-year graduate men tested at four eastern universities in 1941. This change in the characteristics of the groups being tested was one reason for the development of a new scale for the Aptitude Test. A second reason was the fact that the Aptitude Test itself had undergone revisions in content and scope since the original scaling. It might be noted that the Aptitude Test, as such, was introduced in 1949. Prior to that time, the Verbal and Quantitative scores were obtained as two parts of the Profile Tests.

Aptitude Test scores were first reported on this new scale in the fall of 1952. At the same time, new scales for seventeen of the

¹ Paper presented at the 1953 meeting of the American Psychological Association.

Advanced Tests were also introduced. There were two reasons for re-scaling the Advanced Tests. For one thing, fourteen of them had been extensively revised since the previous scales had been established, and the allotted testing time had been extended from one hour and forty-five minutes to three hours. Secondly, it seemed advisable to make a change in the type of score scale used for the Advanced Tests. The previous scales, established in 1947, had been so set that 500 was the mean scaled score on each particular test for a group of seniors majoring in the field covered by that test, and the standard deviation of their scores was 100. A different standardization group was therefore used for each test, with no adjustment made for group differences in ability. Now it is generally found that some fields of specialization attract more able groups of students, while other fields of specialization attract less able students. Also, some fields attract a more heterogeneous group of students than do others. As a result of this kind of self-selection, the groups of candidates taking different Advanced Tests differ markedly from one another with respect to level and range of ability. In the development of the new Advanced Test score scales, it was decided to design them to reflect these differences in ability as measured by performance on the Aptitude Test. It should be emphasized that both this new type of scale and the 1947 scale preserve the rank order of raw scores on a test and therefore both yield the same percentile ranks for a given group of students.

The data for the re-scaling of the Aptitude and Advanced Tests were collected in the spring of 1952. A single standardization group was chosen. This group consisted of two thousand and ninety-five graduating seniors, six hundred and eighty-six women and one thousand four hundred and nine men, in eleven colleges. The colleges were selected in such a way as to yield a fairly wide range of scores and a level of performance near the average of the groups who normally take the tests. In addition, it was judged that the colleges chosen were widely enough known so that the performance of their students could be evaluated by other test users. It was felt that this knowledge would add to the normative value of the scores themselves. The colleges in the standardization group have been listed in the norms bulletins for the tests.

Each student in the standardization group took both parts of the Aptitude Test—that is, Verbal and Quantitative. The new scale for each part of the Aptitude Test was established by a linear

transformation of the raw scores obtained by the *entire* group of two thousand and ninety-five seniors so as to yield scaled scores with a mean of 500 and a standard deviation of 100. In the scaling of the Advanced Tests, a more complex statistical procedure was employed in order to yield scales which would reflect the performance of the particular Advanced Test group on the Aptitude Test. In addition to taking the Aptitude Test, each of the two thousand and ninety-five seniors also took the Advanced Test in his major field of study. For each Advanced Test subgroup, regression coefficients were determined for predicting Advanced Test scores from Verbal and Quantitative Aptitude Test scores. Estimations were then made of the raw score mean and variance of the *entire standardization group* on each Advanced Test, utilizing the regression coefficients and also the measured differences between the entire group and each Advanced Test subgroup on the two Aptitude Test scores. The equations used in this type of score scaling were developed by Dr. Ledyard R. Tucker of Educational Testing Service (see pages 293-4).

Once estimates of the raw score mean and variance of the entire standardization group on each Advanced Test had been obtained, linear transformations were determined which would result in a scaled score mean of 500 and a standard deviation of 100 on each test for the entire standardization group of two thousand and ninety-five seniors.

Figure 1 shows in graphical form the level and dispersion of scores for each Advanced Test subgroup on the two parts of the Aptitude Test and also on the Advanced Test taken by that group. The total group has, by definition of the new scale, a mean of 500 and a standard deviation of 100 on each part of the Aptitude Test. It can be seen that the Advanced Test subgroups differ rather markedly from one another, and from the total group, in the Aptitude Test scores. For example, as may be seen in Table 1, there is a difference of seven points in the mean Verbal score and sixty-two points in the mean Quantitative score between the Chemistry subgroup and the entire standardization group. It is this kind of difference—the differential performance of each Advanced Test subgroup on the Aptitude Test—which has been taken into account in establishing the Advanced Test scales, and reflects itself in a high mean score (530) for the Chemistry subgroup on the Advanced Test in Chemistry.

TABLE I—SCALED SCORE MEANS AND STANDARD DEVIATIONS ON
APTITUDE AND ADVANCED TESTS, BY
ADVANCED TEST SUBGROUP

Advanced Test Subgroup	N	Aptitude Test- Verbal		Aptitude Test- Quantitative		Advanced Test	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Biology	209	486	94	499	87	495	96
Chemistry	180	507	96	562	99	530	101
Economics	239	476	89	516	99	494	97
Education	180	438	86	434	83	446	93
Engineering	151	454	92	570	86	497	98
French	32	520	92	453	72	533	92
Geology	35	473	92	500	86	488	97
German	10	543	69	495	79	505	99
Government	146	498	93	485	89	496	97
History	181	517	93	468	80	506	97
Literature	239	564	98	463	86	548	99
Mathematics	81	508	93	587	90	542	97
Philosophy	31	563	96	521	90	549	97
Physics	49	531	106	633	79	546	101
Psychology	171	527	94	495	86	512	96
Sociology	127	482	93	447	77	474	96
Spanish	34	529	102	451	83	520	99
Entire stand. group	2,095	500	100	500	100	—	—

Table 1 shows the scaled score means and standard deviations on the Aptitude and Advanced Tests obtained by each of the subgroups—essentially the same data as given in Figure 1, but in numerical form. The fact that the Advanced Test mean and standard deviation for any particular subgroup are above or below 500 and 100 means that the Aptitude Test scores for that group differed correspondingly from the Aptitude Test scores for the entire group, and that some degree of correlation existed between Aptitude and Advanced Test scores for that group. The Advanced Test means range from 446, for the Education Test subgroup, to 549, for the Philosophy Test subgroup. The standard deviations range from 92 to 101. By way of contrast, the 1947 scaling method for the Advanced Tests would have yielded a mean of 500 and a standard deviation of 100 for each of these groups. In the new scales, the 500-point on each Advanced Test is the estimated average score

TABLE II—CORRELATIONS AMONG SCORES FOR
EACH ADVANCED TEST SUBGROUP

Advanced Test Subgroup	N	Correlations			
		Aptitude test verbal vs quan- titative	Aptitude verbal vs advanced test	Aptitude quant. vs advanced test	Multiple R: adv. test vs apt. verbal, quant.
Biology	209	.47	.50	.58	.63
Chemistry	180	.57	.38	.53	.53
Economics	239	.44	.57	.51	.64
Education	180	.57	.74	.50	.75
Engineering	151	.55	.58	.48	.61
French	32	.56	.54	.10	.60
Geology	35	.50	.57	.52	.63
German	10	-.09	.09	-.27	.28
Government	146	.47	.69	.45	.71
History	181	.49	.61	.38	.62
Literature	239	.49	.82	.45	.82
Mathematics	81	.29	.20	.46	.47
Philosophy	31	.49	.75	.22	.77
Physics	49	.50	.62	.44	.64
Psychology	171	.38	.60	.46	.65
Sociology	127	.53	.67	.53	.70
Spanish	34	.66	.31	.12	.33
Entire stand. group	2,095	.39	—	—	—

on each test of the *entire* standardization group, not the obtained average score of any particular subgroup. In the 1947 scales, the 500-point on each Advanced Test was the *obtained* average score for the particular group of seniors majoring in the field of the test.

As indicated above, the extent to which the level and range of Aptitude Test scores for a particular Advanced Test subgroup are reflected in the Advanced Test scale depends upon the correlation of each part of the Aptitude Test with the particular Advanced Test. Table 2 shows these zero-order correlations and the multiple correlations. For the total group of two thousand and ninety-five seniors, the correlation between the two parts of the Aptitude Test was 0.39. All but four of the multiple correlations are above 0.60, although only four are above 0.70. In the case of two Advanced Tests, particularly German and Spanish, the multiple correlations were extremely low—0.28 and 0.33. For those tests, it would be

expected that the scaling is relatively unreliable. In the extreme case, where the multiple correlation is zero, the best estimate of the performance of the total group on the particular Advanced Test is the performance of the subgroup (taking the Advanced Test) itself. Since it is these estimates which are arbitrarily translated to a mean and standard deviation of 500 and 100 respectively, the type of scale resulting from a multiple correlation of zero would be identical to the one previously used for the Graduate Record Examinations. That is, it would be one where selective factors affecting the level and range of general verbal and quantitative ability (as measured by the Aptitude Test) would essentially be disregarded in setting the separate Advanced Test scales, since knowledge of that performance does not aid in the estimation process.

It should be pointed out that in the case of six of the Advanced Tests the numbers of cases for the subgroups were less than fifty. In the case of German this number was only ten. Obviously, then, these scales are extremely unreliable. However, since the primary effort in setting these scales was to determine the performance of a generally representative total standardization group, the distribution of cases among the various major fields of study was accepted as it was found. To add more cases to individual subgroups and thus to change the proportions in the separate major field groups arbitrarily would have meant changing the nature and composition of the total standardization group, which, in turn, yielded the best estimate of the proportions in the populations. At any rate, it was felt that since the unreliable scales involved tests that were taken by relatively few examinees, extensive harm would not be done.

In summarizing the characteristics of the new GRE scales, two main points can be made. First, the new scale for the Aptitude Test incorporates normative data for a group of students which is clearly identified and fairly representative of a large portion of the test users. Second, the new scales for the Advanced Test are adjusted for the differences among the groups in Aptitude Test performance. The scales for all tests are therefore based upon the performance of a single, over-all standardization, or reference group. This type of scale system reduces a major source of possible error in interpretation that was present in the old-type scales

where each test was separately standardized on a different group of students.

Under the old scale system developed in 1947, an individual's Advanced Test score indicated his standing relative to the group of people on whom that particular test was standardized. Since the separate test-groups were different (because of self-selective factors) the scale referents for the various Advanced Tests necessarily differed one from another. Under the present scale system, an individual's Advanced Test score indicates his standing relative to the estimated performance of the entire standardization group. Since this group is constant for all tests, it therefore constitutes a single common referent for all the Advanced Test scales.

SUMMARY

The new scales for the Aptitude and Advanced Tests of the Graduate Record Examinations have been developed with the use of data collected on two thousand and ninety-five second semester seniors in eleven colleges which are generally representative of the schools that normally administer the tests to their students. Scores on the Verbal and Quantitative parts of the Aptitude Test were converted to yield scaled score means of 500 and standard deviations of 100. Scores on the Advanced Tests were adjusted to reflect differences in Aptitude Test scores among the groups taking the separate Advanced Tests. It is considered that this type of score scale would lead to fewer errors of interpretation than the previous scales, in which raw scores on each of the Advanced Tests were converted directly to a scale with a mean of 500 and standard deviation of 100.

EQUATIONS USED FOR THE SCALING OF THE GRE ADVANCED TESTS

Notation:

- v, q = scaled scores on the Verbal and Quantitative parts of the Aptitude Test
- x = raw scores on the Advanced Test
- t = entire standardization group
- s = subgroup taking the particular Advanced Test
- $\bar{M}, \tilde{\sigma}^2$ = estimated mean and variance
- M, σ^2 = obtained mean and variance

b = regression coefficient

C_{vq} = covariance between Verbal and Quantitative Aptitude Test scores

$$\tilde{M}_{x_t} = M_{x_s} + b_{xv \cdot q_s}(M_{v_t} - M_{v_s}) + b_{xq \cdot v_s}(M_{q_t} - M_{q_s})$$

$$\tilde{\sigma}_{x_t}^2 = \sigma_{x_s}^2 + b_{xv \cdot q_s}^2(\sigma_{v_t}^2 - \sigma_{v_s}^2) + b_{xq \cdot v_s}^2(\sigma_{q_t}^2 - \sigma_{q_s}^2)$$

$$+ 2b_{xv \cdot q_s} b_{xq \cdot v_s}(C_{vq_t} - C_{vq_s})$$

Linear Transformation to Obtain Scaled Scores (y):

$$y = Ax + B,$$

where

$$A = \frac{100}{\tilde{\sigma}_{x_t}} \quad \text{and} \quad B = 500 - A\tilde{M}_{x_t}$$

THE LEARNING OF SUBJECT MATTER IN TEACHER-CENTERED AND GROUP- CENTERED CLASSES

GERARD V. HAIGH¹ and WARREN SCHMIDT

The Menninger Clinic and the University of California

A perennial problem for the teacher is that of evaluating teaching methods. A contemporary trend in education involves experimenting with shifting control in the classroom from the teacher to the students as a group. The many variations in the control pattern which have been tried may be loosely classified as teacher-centered and as group-centered. One critical question asked is: "Which approach results in greater student mastery of the subject matter?" Existing studies which bear on this question have yielded conflicting answers. Studies by Asch (1), Husband (8) and Guetzkow, *et al.* (6) show small but significant differences favoring the teacher-centered approach. A study by Faw (5) shows a difference favoring the group-centered approach. Studies by Landsman (9), Smith and Johnson (10), Eglash (4) and Wispé (11) show no significant differences between these two approaches. This conflicting evidence suggests the need for further isolation of the pertinent variables.

One factor confounding the results of the above studies is the almighty importance of "The Grade" to goal-straining American students. If getting a good grade depends on acquisition of knowledge, then this grade motivation may induce the acquisition of knowledge independent of teaching method. In all of the above studies indicating no significant difference between teaching methods, it was necessary for the students to do well in examinations in order to get a good grade. In fact, in all but two, the mark on the final examination or the final grade itself was used as the criterion for the amount of knowledge acquired. The two exceptions (9, 10) used achievement tests but also had course examinations. We can probably infer positive transfer from studying for the regular examinations to performance on the achievement tests.

The one study suggesting superior results for the group-centered

¹ This study was done while the authors were teaching at Springfield College, Springfield 9, Massachusetts.

type of class (5) also used the course examination as a criterion. Since student grades were apparently dependent upon this criterion, the results involve the same ambiguity as those studies reporting no difference.

Only in the study reported by Asch (1) were student grades in the group-centered section *not* determined by examinations. Students were required to determine their own grades and were told that the final examination did not count toward their grade. In this study, the teacher-centered section did significantly better on a final objective examination.

These findings suggest that if students in a group-centered class are not "required to learn" via the grade-examination system, then they do not learn as much information as do students in a teacher-centered class.

However, even if we accept this finding from Asch's study, it does not rule out the possibility that some students may learn more in a teacher-centered class while others may learn more in a group-centered class. The study by Wispé (11) highlights the importance of interaction between student personality and teaching method. In all the foregoing studies, students were arbitrarily assigned to experimental and control groups and the results treated as group data. How would consideration of personality differences affect their findings?

The present study is one more attempt to answer the question as to whether teacher-centered classes or group-centered classes are preferable for the learning of subject matter. It is similar to Asch's study in avoiding the influence of grade-striving by incorporating a self-grading system in the group-centered sections. It pushes beyond the Asch study into the area of student personality differences by assigning students to the two types of sections on the basis of their stated preference. Basically, this study addresses itself to the following question: When students are permitted to choose either a teacher-centered class or a group-centered class and when the students in the group-centered class are not required to submit knowledge to the test of the examination-grade system, will there be a difference between these two types of classes as to the amount of subject matter learned?

At the time of this study (1952), Springfield College had a three-course sequence in psychology, consisting of general, child, and

adolescent psychology, required of all sophomore students. On the first day of the second course, students were presented with the following outline describing two teaching methods:

Assumptions Involved in Two Plans for Teaching

Psychology 2

Plan A

(Teacher-centered)

Plan B

(Group-centered)

Curriculum

The instructor knows the facts which are important for the student to learn.
It is the instructor's job to get the students interested in learning these facts.

The students have questions about the subject.
It is the job of the class (including the instructor) to work together in finding answers to these questions.

Classroom Method

Communication: Most of the time, the instructor talks to the students. Sometimes the students talk to the instructor.

Students listen mostly to the instructor.

Rôle of instructor: Initiates the questions, assigns readings, presents the course material, evaluates students on their learning.

Rôle of student: Learns the course material as presented by instructor. Does the readings as assigned. Gives evidence of understanding on examinations.

Evaluation: Instructor assumes full responsibility for gathering evidence and assigning grade.

Everybody talks with everybody.

Students listen to each other as well as to the instructor.

Moderates group discussion, contributes information as needed, suggests readings when requested, participates as an active group member.

Initiates the questions, contributes his opinion, seeks additional evidence, contributes evidence to group, helps arrive at conclusions, helps others contribute to group.

Students assume responsibility for providing evidence. Instructor and student share responsibility for determining grade.

After general discussion of these two methods, each student filled out a brief questionnaire designed to help him weigh the relative merits of each method. At the end of the questionnaire the student stated which method he preferred for the course. The class was organized so that each student's preference was met. One instructor

taught three teacher-centered sections. Another instructor and two graduate assistants² each led a group-centered section.

Twelve weeks later, on the first day of the third course, students were again asked to choose their preferred method, with freedom to change if they desired. At the end of this third course, all sections were given the Horrocks-Troyer Test (?). The examination was cut in half by using every other question in order to ameliorate the effect of fatigue on students in the teacher-centered sections who had just finished a final examination in the same sitting. All students were informed that this test would not affect their course grade.

POPULATION

The subjects in this study consisted of one hundred and ten students who chose one method and followed it through the two courses. An additional fifty students were involved in the original choosing but do not appear in the study either because they changed methods or because they did not immediately follow up the second course with the third course. As reported previously by Ashmus and Haigh (3), in the original choosing, eighty-one students preferred a teacher-centered class while seventy-nine preferred a group-centered class. Of the students continuing in a class of their original choice, there were fifty-five in each group. As reported previously (3), the groups originally choosing each method displayed no significant differences with respect to age, sex, academic aptitude, previous academic achievement or major in college. It is assumed that this equating persists for the whittled-down groups.

While comparability seems to exist within the two groups with respect to general academic achievement, it might be possible that they differed significantly with respect to knowledge of psychology at the outset of the study. To test for this possible biasing factor, the groups were compared with respect to the marks which the members of each obtained on the mid-term examination in Psychology 1.³ A resultant *T*-score of 0.59 shows no significant dif-

² We wish to thank Mable Ashmus and Richard Wiseman for serving as group leaders.

³ The final grade was not used as a criterion because it was determined largely on the basis of term projects which the students were permitted to submit in any quantity they preferred. Thus, the grades came to be influ-

TABLE I—MEAN SCORES ON OBJECTIVE EXAMINATION OF TEACHER-CENTERED CLASS AND GROUP-CENTERED CLASS

	Teacher-Centered Class	Group-Centered Class
Mean score on objective exam.	25.6	24.7
S.D. of distribution	5.89	6.2

C.R. of the difference $t = 0.78$.

With one hundred and eight degrees of freedom, the level of significance is between 0.4 and 0.5.

ference between the experimental and the control group re knowledge of psychology prior to the experiment.

RESULTS

Table I gives the results of this study.

The difference between these two groups in the learning of subject matter is clearly not significant.

The pattern of results among the six sections was the same as that for the experimental and the control groups. There were no significant differences between any two sections.

DISCUSSION

Of the studies cited in the introduction, this one is directly comparable only with that of Asch in that both involve grade determination in the group-centered sections independent of any examinations and so avoid this influence upon knowledge-gain. In the Asch study, the teacher-centered section did significantly better on the two parts of a final objective test, suggesting more knowledge-gain with this type of teaching method.⁴

The present study suggests that this advantage of the teacher-centered method is nullified when students are permitted to choose the type of method which they prefer. Probably this choice is a reflection of personality factors which also influence the capacity of students to respond to these two types of teaching methods.

enced as much by diligence as by knowledge. Incidentally, the mass of paper accruing discouraged the instructors from ever using this system again.

⁴ Asch also reports "supplementary evaluation data indicate that the nondirective (group-centered) students gained a greater knowledge of diversified subject matter and did more independent reading and thinking about psychology than did the control (teacher-centered) subjects."

(Choice might merely be a reflection of previous experience but this seems to have been ruled out in the present study. Although among those students who had had no experience previously with the group-centered method there was a significant preference for the teacher-centered method, among those who had had experience with both there was no significant preference for either). (3) Wispé's study (11) indicates that personality variables are associated with differences in attitude toward directive and permissive teaching. He found that students whom he described via projective tests as "personally insecure" had particularly unfavorable attitudes toward permissive teaching; those whom he described as "independent" preferred permissive teaching; and those whom he described as "satisfied" reacted favorably to both methods. In a corollary study to the present one, Ashmus, Bedrosian and Wiseman (2) found the following personality differences between students choosing these two methods:

(1) Students choosing a group-centered class are more flexible in problem-solving as measured by the Einstellung Test (Wiseman).

(2) Students choosing a group-centered class were better able to cope with inconsistencies and ambiguities in reproducing stories (Bedrosian).

(3) Students choosing a teacher-centered class were more stereotyped and had less insight in describing their self-picture (Ashmus).

The differences found in the above studies were small; some of them were not statistically significant, but all were consistent. They suggest on the one hand the picture of a student who is flexible in his thinking, able to cope with inconsistencies and ambiguities, understanding and accepting of himself and probably able to enjoy and profit from the give and take of group discussion. On the other hand they suggest the picture of a student who seeks the definite, the concrete, and the ordered, avoiding ambiguity and inconsistency, seeing himself (and life) in stereotyped ways and who is probably able to gain most in satisfaction and knowledge from a teacher-centered class. It is probably in some such vein as this that the results of the present study may best be interpreted.

CONCLUSION

Evidence is presented to show that when students are permitted to choose between a teacher-centered class and a group-centered

class and when the students in the group-centered class are not required to learn subject matter by the examination-grade system, there is no significant difference between these two types of classes in knowledge of subject matter at the end of the term.

REFERENCES

- 1) M. J. Asch, "Nondirective Teaching in Psychology," *Psychological Monographs*, 65: 24, 1951.
- 2) M. Ashmus, A. Bedrosian and R. Wiseman, *Clearness of Perception as Related to Choice of a Student-Structured Class*. Master's thesis, Springfield College, 1952.
- 3) M. Ashmus and G. Haigh, "Some Factors which may be Associated with Student Choice between Directive and Nondirective Classes," *American Psychologist*, 7: 247, 1952 (abstract).
- 4) A. Eglash, "A Group Discussion Method of Teaching Psychology," *Journal of Educational Psychology*, 45: 257-267, 1954.
- 5) V. Faw, "A Psychotherapeutic Method of Teaching Psychology," *American Psychologist*, 4: 104-105, 1949.
- 6) H. Guetzkow, E. L. Kelly and W. J. McKeachie, "An experimental Comparison of Recitation, Discussion, and Tutorial Methods in College Teaching," *Journal of Educational Psychology*, 45: 193-207, 1954.
- 7) J. Horrocks and M. Troyer, *A Test for Knowledge of Facts and Principles in Adolescent Development*. Syracuse University Press, 1946.
- 8) R. W. Husband, "A Statistical Comparison of the Efficacy of Large Lecture Versus Smaller Recitation Sections upon Achievement in General Psychology," *Journal Psychology*, 31: 297-300, 1951.
- 9) T. Landsman, *An Experimental Study of a Student-Centered Learning Method*. Unpublished Ph.D. thesis, Syracuse University, 1950.
- 10) H. C. Smith and D. Johnson, "An Experimental Study of Attitudes and Achievement in the Democratic Classroom," 1951 meetings of A.P.A.
- 11) L. G. Wispé, "Evaluating Section Teaching Methods in the Introductory Course," *Journal of Educational Research*, 45: 161-186, 1951.

A STUDY OF THE FAMILY BACKGROUND OF THE GIFTED

WALTER B. BARBE

University of Chattanooga, Chattanooga, Tennessee

Of what influence has heredity and environment been in the development of the gifted child? Specifically, how does the family background of the gifted child differ from that of the child who is average in intelligence? Unfortunately, it is not possible to determine from such a study as this whether heredity or environment played a larger part in the development of the superior intellect of the subject. But by means of such a study, factors in the development of the gifted may be revealed. In a follow-up study of a group of gifted Negro subjects, Jenkins concluded:

... desirable as it would be to know which of these factors (heredity or environment) has been most potent in the development of our subjects, the writer is unable to present any crucial data on the question. The data relative to heredity are meager, and to some extent, superficial; and while the picture of the cultural background is more complete, even here there are intangible factors which elude objective evaluation (1).

The 456 subjects in this study received an I.Q. of 120 or above on the 1917 Form of the Stanford-Binet. These data were taken from the records of the Psychological Clinic of the Cleveland Board of Education. The range of I.Q. is from 120 to 164, with a mean I.Q. of 130.2. The largest number of subjects were in the 125-129 range (37.3 per cent), while almost sixty-two per cent were between 125 and 135. This placed all of the subjects in about the upper ten per cent (2) of the population of the United States at the time they were tested. A large percentage of the group (forty-four per cent) were in the upper one per cent of the population in intellectual ability as measured by this particular test.

All of the subjects were graduates of the Major Work Program of special classes for public school children in Cleveland, Ohio. The data for this study were obtained from information reported on a five-page printed questionnaire which was distributed to the graduates of the program over the last fifteen years. Of those who received the questionnaire, a return of seventy-seven per cent was received.

RACIAL AND RELIGIOUS BACKGROUND

It is difficult to determine the racial background of a group of subjects. The question arises just how far back the subject should trace his ancestry. With regard to the ancestry of gifted subjects, Hollingworth states:

So few data have been gathered to show the proportion of gifted children in relation to race, that it is perhaps scarcely worth while to discuss the topic except to say that we are ignorant of the facts. We have, however, a few studies of the proportion of gifted in samplings of the various races found at present in the United States (3, pp. 68-69).

In a study by Witty (4), the "racial stock included a preponderance of English, Scotch, German, and Jewish ancestors." Ninety-six per cent of the parents were American-born. These findings are similar to those of Terman and Oden:

The reports on racial origin indicate that, in comparison with the general populations of the cities concerned, there is about a one hundred per cent excess of native-born parentage, a probable excess of Scottish parentage and a deficiency of Italian, Portuguese, Mexican, and Negro ancestry (5).

The racial stock of the subjects in the present study is predominantly German, nearly half (47.6 per cent) of the subjects reported having some degree of German ancestry. The next highest group mentioned, twenty-three per cent, was English. The next most frequent were: Hungarian, 14.7 per cent; Russian, 14.4 per cent; and Polish, 10.2 per cent. Hardly any European country was not at least mentioned.

The fact that the population of Cleveland consists of such a diverse foreign element would tend to make the racial background of the subjects different from the subjects in both Terman and Witty's studies. According to the 1940 Census (6) of the City of Cleveland, the largest foreign-born element, about thirteen per cent, was Polish. Czech, Hungarian and Italian each made up about twelve per cent of the foreign-born population. German, Yugoslavian, English and Russian each represented about seven per cent.

Slightly more than two and one-half per cent of the total sample are Negroes. The percentage of Negroes has risen from less than one per cent in 1938 to nearly five per cent in 1952. According to the census, the per cent of Negroes in Cleveland has risen from eight per cent in 1930 to 9.5 per cent in 1940 (6).

About 11.3 per cent of the group was Catholic, 46.3 Protestant, and 38.8 per cent Jewish. Other groups make up the remaining 3.8 per cent. As Hollingworth reports of New York City (3), children of Catholic parents, many of whom are gifted, are commonly being educated in parochial schools. This explains the rather low percentage of Catholic subjects in Major Work classes. The Jewish group appears to be represented in far greater numbers than its proportional share. This is also true of the gifted in the public schools of New York where Hollingworth reports there is "a marked excess of Jewish parentage (3, p. 70)."

ECONOMIC BACKGROUND

An important phase of a study of the gifted which has not received adequate attention is their socio-economic background. This is difficult to determine and, when done in retrospect, is subject to many errors. The procedure followed in this study was to locate the economic tenth of the census tract in which the subject had lived while he was in public school (7). This gave an indication of the rent and property value of the neighborhood in which the gifted subject had been reared. The results of this phase of the study are presented in Table I.

The economic tenth from which the greatest number of subjects came was the seventh, while the sixth and seventh economic tenths included more than fifty-eight per cent of the subjects. This indicates that the background of the majority of the subjects in the study may accurately be described as "upper middle-class."

TABLE I—THE ECONOMIC STATUS OF 456 GIFTED SUBJECTS

Economic Tenths	Per Cent
Highest	
Ninth	1.1
Eighth	7.9
Seventh	10.3
Sixth	37.1
Fifth	21.1
Fourth	11.0
Third	5.0
Second	3.5
Lowest	2.2
	0.7

ORDER OF BIRTH AND SIZE OF FAMILY

It has long been a popular belief that the gifted child is an only child, or, perhaps, has one sibling. In a study of 253 subjects (8), Goddard reported half as being first-born and three-fourths as being either first- or second-born. Of the first-born, forty-five (about eighteen per cent) were only children.

Hollingworth reported that the gifted child had few siblings. In a study by Cobb and Hollingworth, fifty-seven gifted children averaged less than one sibling each (3). With respect to the order of birth, they found that more than one-half of their subjects were first-born.

In 1940, Terman reported that "the parents of the gifted subjects had produced. . . an average of 3.09 (children) per family. He states that this rate would more than maintain the stock, but "it appears likely that the subjects themselves will not equal the fertility rate of their parents" (5, p. 18).

In the present study two questions were asked to determine the size of the family and the order of birth of the gifted child. About 21.8 per cent of the subjects had no siblings, while 42.6 per cent had only one. Almost twenty per cent had two siblings, and seven per cent three.

About twenty-two per cent of the subjects in this study are only children. This is not as large a number as that found by Cobb and Hollingworth, although their finding that "more than half were first-born" (3, p. 180) is substantiated by the fact that 52.5 per cent of the subjects in this study were first-born. About twenty-nine per cent are second-born, and only 9.3 third-born. The data indicate that in this group the gifted child was the first-born in a family of two children.

PARENT OR GUARDIAN OF GIFTED SUBJECTS

Eighty-seven and a half per cent of the gifted subjects were reared by their own parents. The next largest group, 7.2 per cent, were reared by only their mothers. Two per cent were reared by their own mother and a step-father, and 1.5 per cent by their own father and a stepmother. The father or foster parents each reared 0.9 per cent of the total number of subjects.

Witty (9), in studying one hundred gifted children in Kansas, found that most of their parents were American-born. This was

TABLE II—PERCENTAGE OF PARENTS OF GIFTED SUBJECTS WHO WERE BORN OUTSIDE THE UNITED STATES

Both parents U.S.-born	51.3
One parent foreign-born	21.1
Both parents foreign-born	27.5

not true of the subjects in this study. Slightly less than fifty per cent of the subjects had one or both parents who were foreign-born. This is partially due to the large foreign element in the population of Cleveland.¹ It emphasizes the contribution of the immigrant to the mentally superior groups of the country. Table II presents these data.

Even though more rigid government controls have been placed on immigration, no trend is noted which would indicate that fewer of the subjects have parents who are foreign-born.

EDUCATION OF PARENTS

Hollingsworth states (3) that the educational level of the parents of gifted subjects is far above the average for their generation. "In the majority of the cases where the gifted child has been born since 1915, both parents are graduates of high school, and in far more cases than in the population at large both parents are college graduates (3, p. 180)." Since all of the subjects in the present study were born after 1915, it is interesting to compare Hollingsworth's statements with the data obtained for this group.

Of the fathers of the subjects, 38.4 per cent had a grammar school education or less; 33.6 per cent had a high school education; 8.8 per cent trade or business school; and 19.2 per cent had some college. Of the mothers of the subjects, 32.5 per cent had a grammar school education or less; 42.2 per cent had a high school education; 12.3 per cent trade or business school; and 13.0 per cent had some college. The mothers of the gifted subjects on the average appeared to be slightly better educated than the fathers through the high school and business school levels. However, there were more fathers than mothers who attended college.

MARITAL STATUS OF PARENTS

Terman reported (5) that until 1922, 5.2 per cent of the parents of his gifted group had been divorced and 1.9 per cent were separated.

¹ About 20 per cent of the population of Cleveland were foreign-born.

rated. By 1940, the percentage of divorced and separated parents had risen to 13.9.

The data in the present study are not exactly comparable to the results of Terman's study. The information obtained in this study deals with the marital status of the parents while the subject was in public school. It is perhaps comparable to Terman's 1922 data but is definitely not comparable to his 1940 data.

Eighty-eight per cent of the parents of the subjects were living together while the subject was in public school. About 6.3 per cent were either divorced or separated. This is only slightly higher than the report for Terman's 1922 group and is certainly lower than that of the general population. The remaining five per cent consists of cases where one or both of the parents were deceased.

OCCUPATIONAL LEVEL OF PARENTS OF GIFTED SUBJECTS

The occupations of the parents were listed according to the U. S. census classification. The *Dictionary of Occupational Titles* (10) was used to classify the occupations into seven distinct groups: professional and managerial; clerical and sales; service; agriculture, fishery, forestry, etc.; skilled; semi-skilled; and unskilled. The subjects were asked the title and description of the father's occupation. It was possible to classify all but a few of the occupations listed. Where descriptions were not given and two classifications were possible, the data were omitted. Classifications of the parent's occupation were made for four hundred and thirty-seven subjects. Three of the remaining nineteen were on government pensions, while the rest gave no response to the question at all. The data are presented in Table III.

Hollingsworth reports (3) that more than fifty per cent of the

TABLE III—OCCUPATIONAL LEVEL OF PARENTS OF GIFTED SUBJECTS

	Per Cent
Professional and managerial	40.3
Clerical and sales	22.4
Service	3.7
Agriculture, fishery, forestry, etc.	0.2
Skilled	21.5
Semi-skilled	8.2
Unskilled	3.7

children testing above 140 I.Q. have fathers who are professional men or proprietors. The I.Q.'s of the subjects in the present study are not this high, which may partly explain why only about forty per cent of the parents fall into the professional and managerial group. Hollingworth also states (3) that half of the remaining fathers are in semi-professional and clerical occupations. This corresponds to the clerical and sales group of the U. S. Census Bureau, and the data for this group agree with the data in Hollingworth's study.

The fact that over thirty per cent of the parents are in the laboring class, and about forty per cent of these are semi-skilled or unskilled, is noteworthy. It indicates that while the majority of gifted children do come from parents of higher occupational status, the laboring class also contributes a sizeable number.

SUMMARY

In this study data were presented concerning the composition of the group being studied and their family background.

(1) The range in I.Q. of the four hundred and fifty-six subjects was from 120 to 164 with a mean I.Q. of 130.2. Almost sixty-two per cent of the group were within the 125-135 range.

(2) Slightly less than fifty-two per cent of the subjects are females; slightly more than forty-eight per cent are males.

(3) Of the total samples, only 2.6 per cent are Negroes.

(4) About thirty-nine per cent of the subjects are Jewish. The Jewish group is represented in far greater numbers than the size of this group in the total population would lead one to expect.

(5) The economic tenth of the census tract in which the subjects lived while in public schools was most frequently the sixth and seventh. This would characterize the gifted child as being upper middle-class.

(6) The gifted child appears to be either an only child or first-born in a family of two.

(7) Eighty-seven and one-half per cent of the gifted subjects were reared by their own parents.

(8) Almost fifty per cent of the subjects had one or both parents who were foreign-born. These data indicate that the group studied is quite unlike other studies of gifted groups. Previously, the gifted child was found to have American parentage. The high percentage

of foreign-born in Cleveland (approximately twenty per cent) partially explains these data.

(9) The education of the mothers of gifted subjects is slightly higher than that of the fathers, even though more of the fathers went to college.

(10) Forty per cent of the parents were in the professional and managerial group, 22.5 per cent in the clerical and sales, and thirty per cent in the laboring class.

The subjects in this study come from about average backgrounds with respect to occupational level, educational level and marital adjustment of their parents. Economically the majority of them come from an upper middle-class group.

REFERENCES

- 1) Martin David Jenkins, "A Socio-Psychological Study of Negro Children of Superior Intelligence." Unpublished doctoral dissertation, Graduate School, Northwestern University, Evanston, Illinois, p. 53, June, 1935.
- 2) Merle R. Sumption, *Three Hundred Gifted Children*, p. 6. Yonkers-on-Hudson, New York; World Book Co., 1941.
- 3) Leta A. Hollingworth, *Gifted Children*. New York, The Macmillan Co., 1926.
- 4) Paul A. Witty, "A Study of One Hundred Gifted Children," *Bulletin of Education, University of Kansas*, 2: 8, February, 1930.
- 5) Lewis M. Terman and Melita H. Oden, *The Gifted Child Grows Up*. Stanford, California, Stanford University Press, 1947.
- 6) U. S. Department of Commerce, Bureau of the Census, Sixteenth Census of the United States 1940, *Population*, 11: 712, Part 5, Washington, D. C., U. S. Government Printing Office, 1943.
- 7) Howard Whipple Green, *Census Tract Street Index For Cuyahoga County*, fifth edition. Cleveland Health Council, 1951.
- 8) Henry H. Goddard, *School Training of Gifted Children*, p. 129. Yonkers-on-Hudson, New York, World Book Co., 1928.
- 9) Paul A. Witty, "A Genetic Study of Fifty Gifted Children," In *Intelligence: Its Nature and Nurture*, Thirty-Ninth Yearbook, Part II, National Society for the Study of Education. Chicago, University of Chicago Press, 1940.
- 10) Job Analysis and Information Section, Division of Standards and Research, United States Department of Labor, *Dictionary of Occupational Titles, Part I*, second edition. Washington, D. C., United States Government Printing Office, 1949.

DETERMINING THE RELATIONSHIP BETWEEN INDIVIDUAL INTEREST PROFILES AND OCCUPATIONAL FORMS

ALBERT J. KINGSTON, CLAY E. GEORGE and W. PRICE EWENS¹

The Agricultural and Mechanical College of Texas

One of the real problems which faces vocational counselors who employ the Kuder Preference Record in counseling high school or college students is determining the occupational significance of the obtained interest profile. Although certain counselors are content to utilize the *a priori* suggestions presented in the manual, others have turned to the data on various occupational groups for assistance (5). In each succeeding revision of the inventory's manual, Kuder has presented additional profiles or norms which are the result of various studies. It should be noted, perhaps, that these occupational profiles or norms are based upon studies conducted by different users of the inventory and are not the result of unified, systematic research.

This study represents an attempt to investigate the use of these "occupational profiles" when interpreting the interest profiles of a group of college students.

PROCEDURE

The Kuder Preference Record, Form CH, was administered to a group of white, male students enrolled in the Agricultural and Mechanical College of Texas. Two days later the group was given the following instructions: "List in order of preference the five occupations you would most like to enter after college."

The writers hypothesized that students whose listed vocational choices appeared to be consistent, i.e., tended to reflect a similar pattern of interest, would tend to achieve Kuder interest scores which closely resembled the occupational profile or norm for their chosen profession. Similarly it was reasoned that students whose

¹ Thanks are due to A. E. Denton, R. E. Miller and F. E. McFarland, also of the Agricultural and Mechanical College of Texas, for their help in gathering the data for this study.

stated occupational choices reflected a high degree of inconsistency or discrepancy would achieve Kuder scores which less closely resembled their chosen vocation. In the light of the findings of Strong (6) and the theories of Carter (2) and Bordin (1), the hypothesis seemed worthy of study despite the knowledge that individual deviations and lack of refined measurement might preclude verification.

In order to determine whether the stated occupational choices were consistent or not, three counselors were independently asked to sort the listings. A high degree of consistency was found and only those cases on which there was unanimous agreement were used in further study.² Eighty cases deemed homogeneous or consistent and forty-five cases felt to demonstrate a lack of consistency were obtained. The following typical cases may illustrate the difference between the two groups. G. E. C., a freshman, aged seventeen, enrolled in Basic Engineering, listed his choices as follows: (1) Architectural Engineer, (2) Electrical Engineer, (3) Petroleum Engineer, (4) Civil Engineer. It is readily apparent that the occupational choices of the student cluster around Computational, Scientific and Mechanical interests at a professional level. In contrast, H. R. H., a nineteen-year-old sophomore majoring in Civil Engineering listed his choices as (1) Civil Engineering, (2) Farmer, (3) Carpenter, (4) Nuclear Physicist. H. R. H.'s choices reflect a wide difference in level of occupation as well as a discrepancy in interest areas.

After the stated choices had been sorted into the two categories—consistent or nonconsistent—each category was re-sorted according to the students' major courses of study. As a result, four groups with sufficient numbers for adequate statistical analysis were obtained. These groups included Business Administration, Education, Physical Education and Engineering. Du Mas' coefficient of profile similarity (3) was calculated between Kuder's occupational norms and the Kuder profile of each student (except for the Physical Education group for which no applicable norm was available). By use of the Du Mas technique the degree of similarity between the interest profile for each student and the occupational norm would be ascertained. Close similarity between the profiles

² It should be noted that many students were unable to list five choices of occupation so only the first four choices were utilized in the sorting process.

of those students whose stated occupational choices displayed consistency and the Kuder norms would lend support to the hypothesis.

One difficulty lay in the selection of the appropriate occupational norm. Such factors as suitability and sufficient number had to be considered. For example: the scale "All Professional Engineers" was employed because the writers' experience had shown that many freshmen and sophomores change their specified engineering major after they complete two years of basic engineering. The use of this category also allowed the study of a greater number of students than did the more refined occupational profiles. Similar considerations dictated the use of the other occupational profiles, "Office Managers and Chief Clerks" and "All Secondary School Teachers."

RESULTS

The calculation of coefficients of profile similarity (r_{ps}) between individual student profiles and occupational profiles failed to reveal differences between students whose vocational choices reflected consistent patterns of interest and those which did not. Not only was no difference noted between either of the student groups and the occupational profile, but when both student groups were combined for further comparison, the profile coefficients yielded a median r_{ps} of only 0.08.

Although the analysis of profile similarity failed to reveal significant differences, some distinctions in the interest scores of the student groups could be discerned. One of the most readily apparent differences was the higher Mechanical and Scientific scores achieved by engineering students. This observation led to further speculation about the effectiveness of using profile similarity techniques with the Kuder Preference Record. The Du Mas technique provides a method of determining profile similarity by analysis of the slope of the various scales which comprise the profile. In this study it appeared that at least two of the Kuder scales—Musical and Artistic—were nonfunctional in the measurement of the interests of any of these specific vocational groups. These scales, however, affected the coefficients of profile similarity as much as some of the more pertinent scales.

In order to check this observation, an analysis of variance was made of the scores of the groups: Engineering, Business Adminis-

TABLE I—MEANS AND VARIANCES OF THE GROUPS

Major Scale	Engineering (N = 52)	Physical Edu- cation (N = 42)	Education (N = 17)	Business Ad- ministration (N = 14)	Between Variance	Within Vari- ance	F	P
Outdoor	49.93	50.93	43.24	31.57	1551.69	241.27	6.43	0.001
Mechanical	53.71	42.24	43.24	35.07	1534.51	104.67	14.66	0.001
Computational	31.75	22.31	22.24	30.79	893.32	62.79	14.23	0.001
Scientific	49.89	38.43	37.65	28.79	2157.44	117.77	18.32	0.001
Persuasive	33.70	39.83	39.29	52.57	1342.56	152.70	8.79	0.001
Artistic	25.12	21.95	30.59	20.29	102.68	53.50	1.92	n.s.
Literary	14.19	14.93	18.65	21.57	256.59	66.04	3.89	0.05
Musical	10.85	10.26	13.29	11.79	105.91	40.09	2.64	n.s.
Social Service	36.29	53.17	44.76	40.29	2266.79	148.12	15.30	0.001
Clerical	42.52	41.64	42.65	57.43	1164.78	175.96	6.62	0.001

tration, Education and Physical Education. Table I illustrates these results.

As shown by Table I, neither the Musical nor the Artistic score distinguished among the four groups of students. The Literary scale distinguished slightly (five per cent l.c.). In order to study these relationships further, the means obtained on each Kuder scale by each academic group were compared. Significant differences among the groups were noted. Table II illustrates the various scales which distinguished each academic group.

DISCUSSION

Although the original hypothesis regarding the degree of similarity between students who demonstrated consistency in their occupational choices and the occupational profiles failed to gain support, certain questions regarding the use of occupational profiles were raised. The occupational profiles have been regarded as evidence of the validity of the Kuder Preference Record. In the long run, these profiles may prove of much greater value in counseling than those based upon *a priori* consideration. Some counselors already employ them in vocational counseling. In so doing they may attempt to match, by inspection, a student's interest profile with that of the occupational group. At another time they may be content to match the significantly high scales. In both methods, the question arises regarding how great a degree of deviation be-

TABLE II—KUDER SCALES DIFFERENTIATING THE GROUPS AND *t* VALUES

Low High	Engineering	<i>t</i>	Physical Education	<i>t</i>	Education	<i>t</i>	Business Ad- ministration	<i>t</i>
Engineer- ing			Compu- tational	5.41	Compu- tational	4.20	Outdoor	3.92
			Mechan- ical	5.76	Mechan- ical	4.92	Mechan- ical	5.94
			Scien- tific	5.10	Scien- tific	4.04	Scien- tific	6.45
Physical Education	Persua- sive	2.39					Outdoor	4.04
	Social Service	6.83			Social Service	2.40	Social Service	3.43
							Mechan- ical	2.27
Education							Scien- tific	4.67
	Social Service	2.52					Outdoor	2.06
							Mechan- ical	2.21
Business Adminis- tration							Scien- tific	2.26
	Persua- sive	5.07	Compu- tational	3.47	Compu- tational	2.99		
	Literary	3.00	Persua- sive	3.34	Persua- sive	2.98		
	Clerical	3.74	Literary	2.65				
			Clerical	3.86	Clerical	3.30		

$t = 2.00, p = 0.05, t = 2.60, p = 0.01.$

tween the student's profile and the occupational profile should exist before a proposed vocation be excluded from consideration. Another bothersome question which may arise is whether an excessively high interest score on one scale can compensate for a score which is lower than the norm. The reverse of this question is also equally bothersome. One problem which arises regarding the use of these scales is that they are reported in terms of means and no indication of their standard deviation is included in the Kuder Manual. Despite such inherent weaknesses, a sufficient number of vocational counselors feel uneasy when they interpret

Kuder scales. This has led to considerable interest in the development of some sort of statistical or quantitative method of comparing an individual to a norm occupational profile.

This study tends to demonstrate the inapplicability of profile correlation techniques for use with the Kuder. It would seem that a more productive approach to the problem of profile interpretation might lie in the coding of the profiles in a manner similar to that used in the MMPI. This procedure has been suggested by Wiener (7) and elaborated upon by Frandsen (4).

SUMMARY AND CONCLUSIONS

(1) The Kuder Preference Record was administered to a group of male college students who also listed their choices of occupation in order of preference.

(2) Three trained counselors independently sorted the stated choices into two groups; those which demonstrated a consistent pattern of interest and those which revealed an inconsistent pattern. The groups were next divided according to the student's course of study.

(3) The Kuder profile of each student was compared with the most suitable occupational profile in the Kuder manual. The Du Mas coefficient technique for determining profile similarity was used for this purpose. No significant differences were observed between the occupational profiles and either consistent or inconsistent groups.

(4) As differences in the Kuder scores of the different academic groups were observed, an analysis of variance was calculated. These results revealed that only a few scales were important in distinguishing each group of majors from the others.

(5) Students enrolled in Engineering were distinguished from those enrolled in the other curricula by significantly higher scores on the Mechanical and Scientific scales. This group also attained significantly higher scores on the Computational scale than did either the Education or Physical Education group.

(6) Physical Education students were distinguished primarily by attaining higher Social Service scores than either the Business Administration or Engineering group. The Physical Education majors also achieved slightly higher Social Service scores than did other Education majors.

(7) Students enrolled in Business Administration attained

higher scores on the Persuasive and Clerical scales than did the other three groups. They also scored significantly higher on the Computational scale than did either the Physical Education or Education group. As a group the business students also tended to score higher on the Literary scale, although this trend is not highly significant.

(8) The group which consisted of students majoring in Education was not distinguished by any consistent pattern. As the students in this group were preparing to teach a wide range of subjects, it seems likely that they represented a more heterogeneous group than did the others.

(9) These findings indicate that the use of statistical methods for determining the coefficient of profile similarity between an individual's interest profile and an occupational profile may result in misinterpretation. It is suggested that coding the various interest scales in the manner of the MMPI may provide an answer. Such a procedure would enable a weighting of those scales which are pertinent and the elimination of those which are not pertinent to a given occupation.

REFERENCES

- 1) E. S. Bordin, "A Theory of Vocational Interests as a Dynamic Phenomena," *Educational and Psychological Measurement*, 3: 49-65, 1943.
- 2) H. D. Carter, "The Development of Vocational Attitudes," *Journal of Consulting Psychology*, 4: 185-191, 1940.
- 3) F. M. Du Mas, "The Coefficient of Profile Similarity," *Journal of Clinical Psychology*, 5: 123-131, 1949.
- 4) A. N. Frandsen, "A Note on Wiener's Coding of Kuder Preference Record Profiles," *Educational and Psychological Measurement*, 12: 137-139, 1952.
- 5) G. F. Kuder, *Examiner Manual for the Kuder Preference Record*, Chicago, Science Research Associates, 1951.
- 6) E. K. Strong, Jr., *Vocational Interests of Men and Women*. Stanford, California, Stanford University Press, 1943.
- 7) D. N. Wiener, "Empirical Occupational Groupings of Kuder Preference Record Profiles," *Educational and Psychological Measurement*, 11: 273-279, 1951.

BOOK REVIEWS

LEOPOLD BELLAK. *The Thematic Apperception Test and The Children's Apperception Test in Clinical Use*. New York: Grune and Stratton, 1954, pp. 282.

This is a book on the TAT and CAT "meant primarily to be of practical use to the student and practitioner of clinical psychology and psychiatry." The volume takes up in eleven chapters and an appendix the following subjects: theoretical foundations for projective testing; clinical use of the TAT; interpretation of the TAT; the TAT pictures as stimuli; clinical and other special problems; character and defenses in the TAT; use of the TAT in psychotherapy; future developments of TAT; the CAT, nature and purpose of the test, history and theory; interpretation of the CAT; a supplement to the CAT, called CAT-S, its purpose, administration, etc.; publications and work in progress on the CAT; tabulations and schemes for CAT and for CAT-S. There is an excellent index, pp. 275-282.

The greatest value of the book probably lies in the detailed discussions of clinical uses, interpretations and special problems in relation to a large number of cases. This value is found further in the fact that CAT is also presented in similar fashion with some comparisons between TAT and CAT.

Many cases are reported, analyses appear in facsimile printed forms of the Analysis Sheet, and details appear including the descriptive theme, interpretative theme and diagnostic level. Clinical notes follow.

The entire work is the result of thorough preparation and training and long experience in one of the best clinical settings. One cannot go through the details of this volume without becoming well acquainted with much that is being done with the tests. These details stand on their own feet.

It is a question as to whether some of the broad generalizations can be accepted. The experimentally oriented psychologist may look with some wonder at the statement that "The main efforts in American Psychology are directed toward making projective methods the tools of nomothetic science," Whether or not they can agree with this, it does not detract from the mass of clinical material that is of great value for the student of projective testing.

This is a valuable contribution in a special field and in the efforts at progress in the stream of American psychology.

A. S. EDWARDS

The University of Georgia

LEONARD KORNBERG. *A Class for Disturbed Children: A Case Study and Its Meaning for Education*. New York: Teachers College, Columbia University, 1955, pp. 157.

The work of special classes for children who do not fit into the regular school work finds contribution in this study, which is advertised as a book for all teachers. It is "a case study of a class" in a residential institution. The data were gathered during a five-month period.

The fifteen children, all boys, had failed in other schools. "I was met in my classroom with brooding fear and hate." This was "a residential school for disturbed children, all away from home, all under psychiatric care." The author states that out of his experience came "the insight for a basic thing, a truth to renew and care for in teaching." What was done he believes was a "renewal of the oldest form of teaching—teaching by dialogue" (p. 2). This was a dialogue of "persons," not of "words."

Chapter 1 is Teaching as Dialogue, and Chapter 2, The School's Community. The major part of the book lies in the details of Chapter 3, entitled The Classroom History. The three later chapters explain Methodology and Problems, Some Practical Implications, and Toward General Principles.

The details of the classroom history indicate how clarity came in the method of teaching; these are the bases for the theory worked out in the later chapters. It seems to be a rediscovery of the best in teaching that has been known and used by the best teachers of many ages. Many readers may discover or rediscover it with the author. It may appear to some that this best kind of teaching does not need the complexity of theory that it now receives but it is heartening to know that it is being rediscovered—perhaps that it has not really been lost.

The report presents what is considered as a good method, worked out in actual practice, and given in detail for those who are interested in the best kind of teaching. It is well worth the consideration of those who wish to think through the problems of teaching and

to learn about one of the ways in which a good deal of success has been reached.

It is a pity that so good a book has been marred by the comment about how "poor" our schools have been "for our day" (p. 134). One is reminded of the common mass criticism which all too frequently falls upon our schools. This error (of false aggregation, according to the logicians) does not affect the main presentation. But one who had visited some of our best schools might as reasonably say, "How excellent have been our schools in our day." Some of our schools are poor, some are mediocre, but also some are excellent.

A brief bibliography of one page indicates books that are recommended by the author. There is no index.

A. S. EDWARDS

The University of Georgia

PUBLICATIONS RECEIVED

- EDWARD E. CURETON. *Proceedings 1954 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, pp. 135 (paper).
- WINIFRED DE KOK. *You and Your Child*. New York: Philosophical Library Inc., 1955, pp. 147. \$3.75.
- DIRECCAO DE VITOR FONTES. *A Crianca Portuguesa*. Lisboa: Este Livro Foi Composto E Impresso Na Grafica Santelmo, Rua De S. Bernardo, 84 (paper).
- PAUL L. DRESSEL. *Evaluation in General Education*. Dubuque, Iowa: Wm. C. Brown Co., 1954, pp. 333 (paper). \$4.00.
- PAUL L. DRESSEL AND LEWIS B. MAYHEW. *Critical Thinking in Social Science*. Dubuque, Iowa: Wm. C. Brown Company, 1954, pp. 36 (paper). \$1.25.
- PAUL L. DRESSEL AND LEWIS B. MAYHEW. *A Handbook for Theme Analysis*. Dubuque, Iowa: Wm. C. Brown Company, 1954, pp. 78 (paper). \$1.50.
- PAUL L. DRESSEL AND LEWIS B. MAYHEW. *Science Reasoning and Understanding*. Dubuque, Iowa: Wm. C. Brown Company, 1954, pp. 223 (paper). \$3.50.
- WALTER N. DUROST AND STELLA S. CENTER. *Durost-Center Word Mastery Test*. Yonkers-on-Hudson, N. Y.: World Book Company (paper).
- HENRY S. DYER AND RICHARD G. KING. *College Board Scores*. College Entrance Examination Board, 1955, pp. 192 (paper). \$1.50.
- RUTH E. ECKERT AND ROBERT J. KELLER, Editors. *A University Looks at Its Program: The Report of the University of Minnesota Bureau of Institutional Research, 1942-52*. Minneapolis: The University of Minnesota Press, 1954, pp. 223. \$4.00.
- Education for Personal and Family Living: A Working Guide for Colleges*. A Project of The American Social Hygiene Association. pp. 145 (paper). \$1.00.
- ALLEN L. EDWARDS. *Edwards Personal Preference Schedule*. New York: The Psychological Corporation, 1954, pp. 18 (paper).
- ALLEN L. EDWARDS. *Statistical Methods for the Behavioral Sciences*. New York: Rinehart & Co., Inc., 1954, pp. 542. \$6.50.
- HENRY EHLERS, EDITOR. *Crucial Issues in Education: An Anthology*. New York: Henry Holt & Co., 1955, pp. 277 (paper). \$2.25.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

October, 1956

Number 6

F SCALE, GAMIN, AND PUBLIC SCHOOL PRINCIPAL BEHAVIOR¹

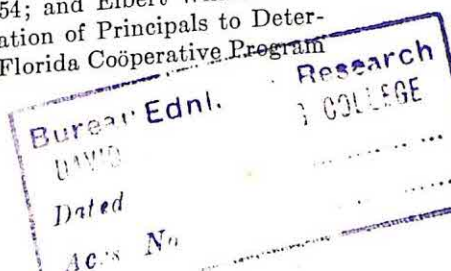
VYNCE A. HINES

College of Education, University of Florida

In the early Iowa studies on influence of leadership patterns upon children's behavior (1), the attempt to train one leader in a democratic operating pattern resulted in the discovery of a new leadership style—the laissez-faire. Although Lippitt and White (2) later reported success in coaching leaders to play three different kinds of rôles, the fact remains that some factor prevented one person from operating democratically. This happy accident served to extend the scope and significance of their research. Did this person have some personality trait or configuration which made it difficult for him to learn the rôle of a democratic leader? Was the cause a failure of training, or was the cause within him?

These questions can be made more general. For some time there has been an emphasis, in our culture, upon the superiority of democratic over authoritarian procedures. Teachers and principals are exhorted to be democratic; besides the philosophic support, assumptions are made that better results can be obtained through using democratic leadership. Hence it could be asked, do some people have a personality structure which makes it difficult or impossible for them to behave in either a democratic or an authoritarian manner? Furthermore, could behavior learned in laboratory or training sessions carry over to the actual work of the teacher or the administrator?

¹Based on data in unpublished research studies by Morton Alpren, "The Development and Validation of an Instrument Used to Ascertain a School Principal's Pattern of Behavior," 1954; and Elbert William Van Aken, "An Analysis of the Methods of Operation of Principals to Determine Working Patterns," 1954. University of Florida Coöperative Program in Educational Administration (CPEA).



Given enough time, sufficient foundation support, and a very large group of people, it would be possible to do extensive initial measuring, differentiated training procedures, and two or three decades of follow-up to see what happens. However, briefer procedures which could provide at least partial answers are also desirable.

There is a wide literature developing on the influence of various kinds of group atmosphere in experimental situations; some work has been done in industry in studying the influence of different leadership patterns upon output and morale; somewhat less has been done in schools.

The present study suggests some partial answers. Data for this report were obtained from a University of Florida CPEA (Co-operative Program in Educational Administration) study on some consequences of different operating patterns of public school principals. Specifically, this study attempts to find possible relationships between the personality factors represented by scores on the F Scale, the Guilford-Martin GAMIN and responses to a principal behavior check list (PBCL).

The F Scale is described fully in Adorno's *The Authoritarian Personality* (3). It was designed to measure an individual's authoritarian tendencies. Items were selected to get at such variables as conventionalism, authoritarianism, superstition, power, cynicism, and projectivity. A high, positive score can be interpreted as representing authoritarian or antidemocratic characteristics in the "inner" structure of personality. A low, negative score would indicate the opposite or absence of these tendencies.

The five factors which make up the Guilford-Martin GAMIN include:

G—general pressure for overt activity.

A—ascendancy in social situations, as opposed to submissiveness; leadership qualities.

M—masculinity of attitudes and interests as opposed to femininity.

I—lack of inferiority feeling; self-confidence.

N—lack of nervous tenseness and irritability.

These tests were selected because it was hoped that they might correlate with behavior as observed in a variety of situations and hence would ultimately prove useful in categorizing or predicting administrative behavior.

To describe the administrative behavior of school principals,

an eighty-six-question check list was developed by Alpren. Criteria for inclusion of questions were that the situation was common to all principals, that the situation involved or affected people besides the principal, that the situation was observable, or observable in theory by others, and that there were a variety of ways of responding to the situation. Situations were collected through discussions with professors of administration and supervision, by conferences with principals, supervisors, and teachers, and by extensive observations in several different types of schools. When the eighty-six situations had been collected, they were divided into three groups and sent as open-ended questions to school principals in various parts of the country. The principals were asked to describe all the ways they used to handle each situation. Their responses were stated in a single sentence and these statements were used as alternatives to the eighty-six situations. From five to fifteen different responses were secured for each situation, each one representing the way at least one principal said he behaved.

The PBCL was used initially in four schools where a research team had worked for a year gathering data through observation and interviews. The principal of each school filled out the PBCL, selecting first, second and third choices for the frequency with which he used various alternatives; each of five teachers, who had been in the school with this principal at least two years, also filled out the check list independently. Another independent check list rating for each principal was based on data gathered by the research team. Using rigorous criteria for correspondence, about eighty-five per cent agreement was found between the research team description and a composite of teachers' descriptions; about seventy per cent correspondence between team and principal; and slightly less than seventy per cent correspondence between teacher composite and principal.

With some minor revisions, the check list on principal behavior was then filled out by sixty white and fifteen Negro principals, plus five teachers in each of these seventy-five schools. The teachers had worked with the principals they rated for at least two years. Principals and teachers were asked to name first, second and third choices for the frequency with which principals responded to the situations. The check lists were completed by the principals on one occasion; by the teachers on another. In general, all respondents worked independently.

The principals also completed the F Scale and the GAMIN

just prior to working on the PBCL. Other data on the principals included sex; length of service as a principal in the county and in the particular school; highest certificate held (Florida gives a Rank III Certificate for four years of college, Rank II for five years, and Rank I for six years or more, and while these extra years are often in education they can be in content fields such as political science, chemistry, botany, etc.); responses to an open-ended situational questionnaire; and, for part of the principals, responses to two other personality tests. These last three proved not to be helpful and were abandoned.

Responses to the PBCL were placed on IBM cards and weighted composite responses were determined for each principal. A first choice counted three, second choice two, and third choice one. From lack of internal consistency and lack of agreement among two white principals and their respective five teachers, results from these schools were not used in later calculations.

If perfect agreement existed among a given principal and the five teachers from his school on a first choice, this would receive a composite total of eighteen points; a second choice, twelve points; a third choice, six points. This, of course, seldom happened. However, in most instances two choices usually stood out from others. In determining the principal's most probable behavior two choices were finally used, and no distinction was made between them. These two were arrived at by (1) agreement between the principal and three or more teachers; (2) agreement among more than three teachers; (3) where neither (1) nor (2) held, the principal's statement was accepted as the most probable correct description of his behavior. This latter practice was based upon extensive rechecking in six white and five Negro schools to try to resolve indeterminate situations and by an examination of agreements of responses of various teachers and the principal with composite responses in six schools of each race, the schools having been selected with random numbers. Items in which neither (1) nor (2) held in more than half the schools were deleted.

After the two most probable responses had been determined for each principal for the usable situations, each principal's responses were cut as a scoring key. The original plan for the research had been to find groups of principals who responded the same way to many of the key situations and then to examine the bases for such similarities. Various mechanical means were used to try to isolate

groups with similar patterns. These did not work. External criteria were used singly and in combinations to separate out groups with similar patterns. These included personality factors, experience, educational level, and size of school. These were unsuccessful. At this point the research team reluctantly decided to try to categorize behavior as democratic, authoritarian, and laissez-faire. A team of three faculty members who had worked together for several semesters in a leadership program for principals and supervisors agreed to attempt the classification. They represented administration, elementary education, and secondary education. They developed and unanimously agreed upon the following criteria for behavior:

(1) Democratic behavior.

- (a) Action involving the group in decision making with respect to policy and program.
- (b) Implementation in line with democratically determined policy.
- (c) Action promoting group or individual creativity, productivity.
- (d) Behavior or attitude respecting the dignity of individuals or groups.
- (e) The principal seeks to become an accepted member of the group.
- (f) The principal seeks to keep channels of communication open.

(2) Authoritarian behavior. Authoritarian behavior generally was defined as the opposite of behavior described in (a) to (f) above. Specifically, authoritarian behavior was defined as:

- (a) Decision-making centered in the status leader or his inner circle.
- (b) Obtains objectives by pressures that jeopardize a person's security.

(3) Laissez-faire behavior—taking no action, shirking responsibility, “passing the buck.”

(4) Indeterminate behavior—any behavior on which the three judges could not unanimously agree was called indeterminate.

Thirty-one situations were eliminated because the judges did not find both authoritarian and democratic responses among the alternatives. Keys were prepared using the classified responses. All the democratic options were cut on one key, authoritarian

responses on another, laissez-faire on a third, and indeterminate on a fourth. Thus a number could be obtained for the frequency of each classification of behavior for each principal. Authoritarian responses varied from 20 to 46 and democratic responses from 24 to 61 for white principals and 20 to 38 and 23 to 44, respectively, for Negro principals. Laissez-faire items varied from 0 to 7 and indeterminate responses from 19 to 31 for the two groups. Because of the small spread and low frequency of laissez-faire behavior, it was not used in later procedures.

A second jury of ten professors who had had administrative experience in public schools also classified the items. Each person in this group, operating independently, picked out what he considered the two most effective responses to each situation. A composite was made of these answers and the principals were again ranked according to frequency of use of what the experts called "most effective" practices. These varied from 46 to 82 for white principals and 46 to 70 for Negro principals. As Table I indicates, scores for both groups were fairly similar and well distributed.

Hence it was possible to explore relationships among scores on the F Scale and the GAMIN with frequency of democratic behavior, of authoritarian behavior, and of use of most effective practices. Table II shows the correlations obtained. An examination of Table II shows that there are no significant correlations between the six personality factors used and the frequency of three different classifications of behavior.

The only variable investigated in the study which seemed to correlate with frequency of behavior was sex. Women principals were significantly more often democratic, less often autocratic,

TABLE I—RANGE, STANDARD DEVIATIONS, AND MEANS FOR 58 WHITE PRINCIPALS AND 15 NEGRO PRINCIPALS ON SIX PERSONALITY FACTORS

Range s \bar{X}	F Scale		Guilford-Martin									
	White	Negro	G		A		M		I		N	
			White	Negro	White	Negro	White	Negro	White	Negro	White	Negro
	110	60	22	11	22	15	30	21	34	12	27	19
	22.5	20.1	4.7	3.8	5.0	4.7	6.7	5.1	6.4	3.4	5.9	5.4
	-5.2	7.8	10.7	10.8	21.2	23.9	19.3	22.9	38.1	39.5	33.2	33.1

TABLE II—CORRELATIONS BETWEEN SIX PERSONALITY FACTORS AND FREQUENCY OF AUTHORITARIAN BEHAVIOR, DEMOCRATIC BEHAVIOR, AND MOST EFFECTIVE PRACTICES FOR 58 PUBLIC SCHOOL PRINCIPALS

	F Scale	Guilford-Martin				
		G	A	M	I	N
Frequency of authoritarian behavior	-0.146	0.029	0.214	0.506*	0.052	0.044
Frequency of democratic behavior	-0.212	0.046	0.099	-0.410*	0.002	0.078
Frequency of most effective practices	-0.318	-0.062	-0.235	-0.516*	-0.254	-0.259

* Correlations were spuriously high for masculinity scale because of bimodal distribution of M scores. Naturally the men public school principals in the sample were found to be much more masculine than the women principals. The correlation between authoritarian frequency and the M factor for males was 0.041; for females, 0.164. There were thirty-two male and twenty-six female principals in the sample. N was 58 for the above zero-order coefficients. Other correlations disappeared when calculated by sex. Similar results were found for the Negro principals.

TABLE III—MEAN FREQUENCY OF WHITE MEN AND WHITE WOMEN PRINCIPALS ON DEMOCRATIC BEHAVIOR, AUTHORITARIAN BEHAVIOR, AND MOST EFFECTIVE BEHAVIOR

	Authoritarian Behavior	Democratic Behavior	Most Effective Behavior
Men	37.0	35.9	58.8
Women	29.7	43.8	69.3

and more often users of best practices. There was a point biserial correlation of 0.595 between sex membership and frequency of most effective practices. For frequency of democratic behavior the correlation was 0.579; for authoritarian behavior, the correlation was 0.588.

Hence it can be concluded that the personality factors measured by the F Scale and the Guilford-Martin GAMIN do not predict frequency of democratic or authoritarian behavior or use of best practices on the job. This lack of relationship between these personality measures and frequency of different types of behavior would suggest that principals could learn to use any of these operational patterns of behavior.

REFERENCES

- 1) R. Lippitt, "An Analysis of Group Reactions of Three Experimentally Created Social Climates," Unpublished doctoral dissertation, State University of Iowa, 1940.
- 2) R. Lippitt and R. White, "The 'Social Climate' of Children's Groups," In R. C. Barker, J. Kounin and H. Wright, Editors, *Child Behavior and Development*, pp. 485-508. New York, McGraw-Hill, 1943.
- 3) T. W. Adorno et al., *The Authoritarian Personality*. New York, Harper & Brothers, 1950.

ERRATUM

In "Effects of Non-promotion on Educational Achievement in the Elementary School," by William H. Coffield and Paul Blommers, Vol. 47, No. 4, April, 1956, p. 249, first sentence under the section entitled "Concluding Remarks" should contain the words "does not appear" rather than "does appear."

RELATIONSHIP OF SUCCESS IN LOWER AND HIGHER LEVEL MILITARY SCHOOLS

JAMES H. L. ROACH

Tulane University

The many studies on civilian education have shown a high positive relationship between success in high school and college and between success in college and in professional schools. Little has been reported thus far about the relationship between success in lower level and higher level military schools.

To determine if any relationship existed, a study was made of three hundred and nineteen students in the Air Command and Staff School (AC & SS), classes 51-C, 52-A and -B, and 53-A, who had previously attended the Air Tactical School (ATS).¹ The relationship between four "recommendation" groups of ATS graduates and their final cumulative weighted mean t-scores in AC & SS is as follows: Group 1 recommended for further schooling after ATS (N=111), 56.43; Group 2 tentatively recommended for further schooling after ATS (N=99), 52.08; Group 3 tentatively not recommended for further schooling after ATS (N=60), 48.77; Group 4 not recommended for further schooling after ATS (N=49), 46.22. The differences between each of the four recommendation categories are statistically significant.

In addition to the above data, a correlation coefficient of 0.68 was found to exist between grades of two hundred and thirty-nine students who had attended both schools. The smaller number of students used here is due to the fact that ATS grades were not available for some students for whom there were recommendations.

¹ In the Air Force educational program the ATS (now the Squadron Officers School) is designed for lieutenants and captains. The AC & SS (now the Command and Staff School) is designed for majors and lieutenant colonels. Both are a part of the Air Command and Staff College. The highest level school in the Air Force (for colonels) is the Air War College.

GROUP THERAPY AND REMEDIAL READING

AARON LIPTON¹ and ARTHUR H. FEINER²

The purpose of this paper is to describe one type of group therapy approach to the problem of reading retardation. From March 1955 to June 1955, a group of fourth grade boys were referred to the authors because of failure in all academic subjects, particularly reading. The authors were consultants, at that time, to a Westchester County, N. Y., public elementary school. Attempts had been made to provide these boys with a remedial reading experience, but because of their own personality disturbance this proved entirely ineffective. In this situation the boys had been presented with material that was anxiety provoking in that the act of remedial reading reminded them of their original deficiencies and lowered self-esteem, and yet had been permissive in allowing them to act out their harbored resentments toward the school and each other. The situation was entirely too frustrating as well, so that no growth was possible. The answer for this group lay not in forcing controls on the children since this had been done for four years without much learning taking place.

If the children were to learn to read, some other approach was indicated. The new direction took cognizance of the fact that reading is used at times as part of the security operations of the individual. Reading as a cultural tool is dependent on the psychological growth and maturity of the child. "The failure or lack of desire to read is a reflection of an inability to adopt more mature patterns of social relatedness" (1). Reading experiences can be part of the child's difficulties in living in the sense that they may be involved directly in the child's own conflicting problems of adjustment. Thus, failure in reading can be used as an aggressive weapon against an authority figure, parent or otherwise, or against oneself. An attempt to describe the precise nature of these aspects of the problem has been summarized by others (2, 4). The crucial point is that reading can be learned only when the child is ready. In view of these assumptions the direction had to be one that would eliminate the use of reading as an avenue of negativism.

¹ Reading Consultant, Public Schools, Greenburgh, N. Y., and Instructor in Reading, New York University.

² Psychotherapist, Blueuer Psychotherapy Center, Jamaica, N. Y.

Group psychotherapy was decided upon and the group met three times a week for hourly sessions. The focus was an activity program designed to meet the needs of the children.

The group consisted of nine fourth-graders, all of average or above average intelligence. They were from ten to eleven years of age. All were from markedly deprived socio-economic levels and all manifested patterns of disturbed interpersonal living and academic adjustment. Five of the boys were characterized as outwardly aggressive, often destructive in their behavior. Four were predominantly timid and withdrawn. One of the "aggressives" and one of the "timids" were considered "isolates."

In a previous publication (1) the concept of group development within the framework of emergent levels was presented. A brief recapitulation of these ideas follows.

The group's development went through several stages of growth, from an anarchical, amorphous, disorganized unit to an integrated, verbally functioning unit. The group progressed through the following stages or levels, none of which is all-inclusive or exclusive of each other: (1) disorganization; (2) level of individual competition (each boy competes with each other boy); (3) level of team competition (goal-oriented team play); (4) level of group solidarity and identity (characterized by phrases "we should," "let's all," "why don't we"); (5) level of verbal interaction that is socially adaptive (jokes, sharing ideas, experiences and personal information); and (6) level of self-directed learning activity ("we don't want to play, we came here to read!").

Reading gains were measured by the Smith "Informal Reading Inventory."³ Table I summarizes the gains for each child. Since little gain is normally expected in a four-month period, the gains these boys made would appear to be quite significant.

Social gains were equally marked. From the start of group therapy to the termination of the program there was a steady diminishing of hostile and/or withdrawn behavior, increasing social awareness and interaction, and an increase in socially accepted goal-directed activity.

An explanation of the gains is in order. The goal in the treatment program was fixed at bringing the children to the level of reading readiness in a complete motivational and maturational

³ N. B. Smith, "Informal Reading Inventory," unpublished.

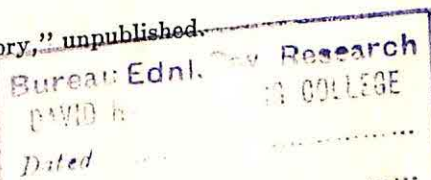


TABLE I—GAINS IN READING LEVEL

Name	Reading Levels	
	Pre-therapy	Post-therapy
M.M.	Pre-primer	Pre-primer
A. L.	Pre-primer	1st grade—1st half
P. R.	Pre-primer	1st grade—2nd half
N. L.	Pre-primer	1st grade—2nd half
V. R.	1st grade—2nd half	3rd grade—1st half
R. O.	1st grade—2nd half	3rd grade—1st half
E. D.	1st grade—2nd half	3rd grade—1st half
M. R.	1st grade—2nd half	3rd grade—1st half
J. M.	2nd grade—1st half	3rd grade—2nd half

sense. Since children do not read for reading's sake (1), skill instruction, ill-timed or premature, is fraught with danger and often only serves to further complicate the process. If the avoidance of reading is integrated by the child as a defensive gesture, the reduction of this need enables the child to incorporate further learning as part of social living. By therapeutically attacking the basic causes for this mal-adaptive integration, the children emerge at a true level of readiness. The goals in all therapy are essentially those that "offer the patient the means of reestablishing feeling-interest relationships with his social environment . . . [and] . . . the restoration of the patient to full membership of society . . ." (3, p. 213). By offering interpersonal situations that involved the unqualified acceptance of each boy's feelings, the recognition of each boy's needs, and the consistent offering of the therapists' own selves as helpful loving adults, the original destructive and isolating techniques necessary for defense of a basic need to survive could be replaced. This is not to say that the therapists maintained an all-permissive position of their own in their attempts to be loving. Those aspects of behavior which were either personally or socially (group) destructive were rejected with redirection.

These comments do not negate or deny the important need for skill instruction. Skill instruction must interpenetrate the need-level of the child and cannot be superimposed by mechanical, additive conceptualizations. Its place and timing must be determined by teacher, therapist or consultant with need-level as a consistent frame of reference.

Another aspect to consider is the predominant problem of reading in relation to other learning processes. For example, many more problems emerge in reading than arithmetic. One explanation for this is that children's needs to read are much more dependent on socially-oriented goals, while a need to learn arithmetic is much more basic to survival in our culture. (Reading does become more needful towards adulthood.) Handling numbers occurs very early in years. For example, buying candy, milk in school, getting an allowance, paying carfare, purchasing, counting toys and games are all experiences which bring arithmetic closer to the immediate needs of the child. On the other hand, reading is not nearly as vital a part of a child's life. One could more conceivably do without reading until a much later age. In addition, the need for a child to read is often much more the parent's need. Thus the parent, who is so involved in all other aspects of a child's development, may become excessively involved in imposing an adult-level need, such as reading, on the child before the child is ready. This concern may be picked up by the child and integrated by him as a reflection of his feelings of inadequacy. When a parent says to a child, "Why don't you read better?" he may be saying that he doesn't think his child is as good or smart as the child should be. Reading then can easily become a potent weapon in the hands of the child. For if the child learns that reading is too important to the parent he may avoid reading as one expression of hostility that is still available to him.

More boys than girls are victims of this kind of interaction because the boys in our culture ultimately must become breadwinners and parents are legitimately concerned that their children be "successful." The feeling that "girls will get married anyway" may mean that the girl will not be bothered by her parents and thus may evade the need to use reading as a hostile operation. Obviously there are other factors which affect children's reading progress, but underlying the problem generally is the lack of consideration of focusing the teaching of anything at the needs-level of the child. The implications of these remarks transcend working with disturbed children. They apply to the development of curriculum generally, with well-adjusted children who also have needs-levels which must be considered. The idea of gratifying the child's needs is certainly not new but it is time for a reformula-

tion of educational practices and theory in terms of emergent readiness.

REFERENCES

- 1) Arthur H. Feiner, and Aaron Lipton, "Group Therapy as an Approach to Improvement in Reading," *Journal of Child Psychiatry*, August, 1952.
- 2) Gerald H. J. Pearson, "A Survey of Learning Difficulties in Children," in *Psychoanalytic Study of the Child*, Vol. VII, pp. 322-386. New York, International Universities Press, 1952.
- 3) Ian D. Suttie, *The Origins of Love and Hate*, 275 pp. New York, Julian Press, 1952.
- 4) Edith Weiskopf, "Intellectual Malfunctioning and Personality," *Journal of Abnormal and Social Psychology*, 46: 410-423, July, 1951.

BOYS WHO RUN AWAY FROM HOME

WILLIAM W. WATTENBERG

Wayne University

Truancy from school is known to vary in its significance as a symptom. On the one hand we have the "circus" truanacies and "swimming hole" truanacies, which appear to represent an effort by relatively "normal" boys to add a bit of variety to the drab routine of day-to-day existence. On the other hand, we have the escape from school of the youngster whose pressing emotional problems will later find outlet in serious delinquency.

Truancy from home includes during the regular school year truancy from school as well. On the surface at least it would appear to be even more significant. To the extent that it represents escape from unsatisfactory living it would appear to mean that the youngster's life is irksome in all areas, not merely in the classroom. It also represents a very drastic break with what should be the young person's chief sources of security—his family. The runaway cuts himself off from a sure supply of food and shelter. In exchange for what?

It is no wonder, then, that to answer so puzzling a question we find much speculation and some research. The theories range from those looking at inward drives to those stressing multiple causation of environmental forces.

One interesting explanation, reminiscent of G. Stanley Hall, was put forward by Staub (8) on the basis of a psychoanalysis. He felt that running away harks back to the childhood of mankind when humanity was in a migratory stage.

Taking a rather different tack, Jahr (4) saw in running away an attempt by young folks to express themselves in an environment over which they had little control. When present surroundings appear burdensome, distant places might seem to offer an opportunity to exercise curiosity, imagination and love for romance or adventure.

Most of the serious fact-gathering took place in relation to the appearance of large groups of wandering youth during the Great Depression of the 1930's. Lovejoy (5) attributed the runaways of that era to the breakdown of relief services and to local failures to

provide vital schooling or adequate recreation. Nylander (6) felt that the trouble was due to failure in school, family disorganization and the breakdown of religious institutions and moral codes. Adding to these lists, Sullenger (9) declared that children leave home because of emotional crises, school problems, too complex a social life and failure to achieve satisfying social relationships.

On the basis of a study of three thousand, three hundred and fifty runaway boys who registered with the Los Angeles Office of the Federal Transient Service in 1934 and 1935, Outland (7) tended to exonerate the schools. He declared that basic economic and social conditions were much more important than educational factors for boys going on the road. Almost half the group had been out of school more than two years before leaving home.

In many ways the most careful studies were the series reported by Armstrong (1-3) on six hundred and sixty boys arraigned in the Children's Court of New York City. Since her group had run away between 1926 and 1930, the influence of economic forces was not as strong as in Outland's. She found that escape from current pressures was the basic motive in most cases. As compared with boys involved in theft or charged with incorrigibility, the runaways were younger and had higher IQ's. The runaways came largely from homes of low economic level, where the mother was working. More than half came from disrupted families. The group seemed to be high in its proportion of only children. There was also a marked tendency for oldest children to be runaways. Seventy per cent of the group said they left because of trouble at home; thirty-seven per cent mentioned trouble at school. Armstrong regarded running away as a psychoneurotic reaction in which irresponsible or unfit parenthood was the basic cause and the precipitating events might arise from either school or family factors.

PROCEDURE

The present study is based on records of five hundred and seventy-five cases of runaway boys dealt with by the Crime Prevention Bureau of the Detroit Police Department.¹ When-

¹ Acknowledgment is due a large number of individuals who aided in this study. The procedures were inaugurated by Inspector Sanford Shoults, the inquiry form was developed by Inspector Ralph Baker, the statistical work was performed by Lieutenant Francis Davey and the data were gathered by the patrolmen assigned to the Bureau.

ever a boy was reported missing, a trained police officer visited his home to gather information, which included his parents' ideas as to why he had run away. When the boy was found or after he had returned home, he was interviewed as to his reasons for leaving home. As a final measure, the boy was again interviewed in a family conference setting. This not only served to produce better family planning for the future, but the interplay of explanations added to the officer's understanding of the forces leading to the truancy. All the data thus gathered were recorded on a thirty-item checklist.

It should be pointed out that the population studied was quite different from those used in the studies previously reported. In contrast to the Outland population, this Detroit group was investigated during a period of relative prosperity and high productivity. In contrast to Armstrong's cases, few of the Detroit group were delinquent in the official sense. These were not court cases but rather a group reported as "missing." In the vast majority of instances, no official charge was made. The parents were happy to have their sons back, the boys were relieved to be home again, and no official action was taken against anyone.

FINDINGS

The peak chronological age of the runaway boys was fifteen. Of the five hundred and seventy-five, the fifteen-year-olds accounted for one hundred and sixty-nine. The age distribution of the entire group is set forth in Table I. It would appear that truanting from home is an adolescent rather than a pre-adolescent phenomenon.

The matter of broken homes was checked by two different techniques. One was to inquire regarding with whom the boy was

TABLE I—AGE OF RUNAWAY BOYS

Age	Number
	24
10	29
11	49
12	82
13	109
14	169
15	110
16	3
17	

TABLE II—WITH WHOM BOY WAS LIVING WHEN HE RAN AWAY

Living with:	Number
Both parents	330
Mother	83
Father	24
Father and stepmother	42
Mother and stepfather	41
Grandparents	12
Uncle and aunt	6
Brother	4
Sister	2
Cousin	3
Rooming-house owner	1
Foster parents	4
Boarding home couple	11
Institutions	3

living; the other to record the marital status of the parents. Of the five hundred and seventy-five boys, three hundred and thirty were recorded as living with both parents. In eighty-three cases, there was a step-parent in the home; nineteen boys had run away from institutions, boarding homes or foster parents. In the remaining cases, the boys came from households where there was only one parent present, or where there were relatives. The detailed story is given in Table II.

The reports as to the marital status of parents presented a substantially similar picture, although there were discrepancies due to the ways in which the questions were interpreted. The main point is that slightly more than three hundred of the homes were recorded as intact; approximately two hundred and fifty had been broken in one way or another. The details are presented in Table III.

An effort was made to check Armstrong's findings relative to sibling relationships. Only children accounted for forty-four of the cases. Census figures indicated that approximately ten per cent of Detroit children had no siblings. On this score Armstrong's findings were contradicted by the Detroit group. Of those with siblings, one hundred and seventy-five were the oldest; one hundred and thirty-two, the youngest; and two hundred and twenty-four occupied middle positions. The slight preponderance of first-born as compared to last-born would bear out that portion of her con-

TABLE III—MARITAL STATUS OF PARENTS

Status	Number
Real parents living together	310
Common-law relationship	2
Mother remarried	71
Father remarried	64
Parents separated	70
Father dead	30
Mother dead	14
Not stated	14

clusions. Sex of siblings was not a significant factor: three hundred and eighty-six boys came from homes where there were one or more brothers; three hundred and eighty, where there were one or more sisters. Where parents had remarried, there was a slight tendency for running away to be linked to the presence of stepsisters: forty-three boys came from such homes, as compared with thirty-six from homes where there were stepbrothers.

The matter of deciding the reason for the running away was not easy. In most cases, as the evidence unfolded it became apparent that there was an interaction of several factors. However, an effort was made to determine the chief or precipitating cause for the incident itself. On this basis, it was the judgment of the investigating officer that search for adventure was immediately responsible for one hundred and twenty-four of the truancies. Rebellion against parents figured as the main cause for one hundred and fifteen boys. Escape from school conditions led to eighty-seven incidents; and fear of punishment for some misdeed appeared in seventy-six cases. The truancy was part of an attempt to evade the police in thirty-two cases. The remaining boys gave a wide range of reasons.

School conditions did not appear as either a major or a contributing factor in two hundred and fifty cases. Where school was involved at all, the trouble was likely to be attributed to poor marks; these were mentioned by one hundred and eighty-one of the boys. Twenty-five boys complained that they were unable to keep up with the class; twenty-eight felt teachers were persecuting them; and eight left home to escape an examination. Seven boys felt they were objects of ridicule by classmates; seven were swamped by being in too large classes.

As to home conditions, these were found not to blame in two

TABLE IV—HOME CONDITIONS MENTIONED AS CAUSES
FOR RUNNING AWAY

Condition	Number
Broken home	42
Parents too lax	42
Parents too strict	27
Nagging by members of family	27
Home too crowded	26
Blocking of boy's plans and desires	16
Ridicule by members of family	15
Neglect by parents	12
Home too dirty	12
Siblings pick on boy	12
Parents too bossy	10
Parents pick on boy	9
Stepmother picks on boy	9
Stepfather picks on boy	8
Parents too old-fashioned	8
Mistreatment due to drunkenness	9
Parents baby boy too much	6
Too many chores	5
Parents refused spending money	5
Friends not permitted in home	5
Boy not allowed to drive car	4
Jealousy of other family members	5
Immorality in home	2
Boy not allowed to have pet	1

hundred and fifty-seven cases. Where they were mentioned, there was a wide range, as indicated in Table IV. Linked to most such conditions was rebellion against severe treatment or, by contrast, distaste for laxity.

As previously mentioned, the search for adventure figured as a main reason in one hundred and twenty-four cases. It was mentioned one way or another in two hundred and sixty-seven interviews. In one hundred and twenty cases, travel was described as being alluring in its own right; in one hundred cases the adventurous element was too vague to pin down more definitely. For sixteen boys, an exciting job was the goal. Fourteen ran away to go camping; eleven, to go fishing; three, to go hunting; and three, to join the armed services.

In a number of cases, the abrupt departure from home was motivated by the need to escape the aftermaths of some other

delinquency. In twenty-three cases, this was an auto theft; in eighteen cases, a larceny; in seventeen cases, a burglary; in the remaining twenty-six instances, a miscellaneous collection of misdeeds.

Not too easy to classify were the other reasons brought out in the interview and conferences. A few of these merit mention. For seventy-two of the boys, the running away was an effort to spite someone; indeed, a number were found hiding where they could witness and revel in the anxiety created by their absence. In sixty-four cases, part of the picture was desire by the boy to keep the good opinion of a companion who had his own reasons for taking to the road. For forty-eight boys, the destination was the home of some relative whom they felt would treat them better than their parents. Employment was the goal of forty-three. Another thirty-seven were influenced by a bad gang. For twenty-seven boys, the running away was intended to provide proof of self-sufficiency.

At the time of the initial contact, when the boy was missing, the officers rated the attitudes of the parents. With few exceptions, these were coöperative. Of the entire group, only nineteen parents could be described as indifferent; four were secretive; one, arrogant; and one, antagonistic.

After the boy was recovered, rather few parents changed their tunes. The police officers described two hundred and thirty-seven as forgiving; one hundred and sixteen, understanding; seventy-three, as seeking assistance in doing better with their sons; and forty-eight, as intent on showing affection. On the less favorable side, forty-three turned punitive; eighteen became openly rejecting; eighteen were over-protective; and three refused assistance.

The boys showed an equal range of reactions upon their return: two hundred and seventy-eight were repentant; one hundred and sixty-three, indifferent; forty-five, sullen; twenty-seven, defiant; twelve, fearful; and twelve, boastful. The remaining thirty-eight were harder to describe in a single adjective.

DISCUSSION

It must be stressed that the five hundred and seventy-five boys under study probably represent a fair sample of all boys who run away from home. They became known to the police not because of complaints, but because their parents were worried and wanted them back. Few were delinquent in the official or legal sense of

that term. For most, no complaint was filed. Rather, the incident led to a rallying of family forces which seemed to portend a happy outcome. In such cases the experience of the Youth Bureau has been that less than twenty per cent ever came again to the attention of police authorities.

A very striking fact is the large proportion of the group for whom the main motivation would seem to be a search for adventure, an effort to exercise independence. For this portion of the sample, the truancy seems to have had a positive significance. It appears as an index of striving, perhaps ineptly, for a more adult status.

More concern must be given to those instances where poor home or school conditions had reached the point where the youngster found the situation intolerable, and his personality organization was such that he fled from it. Here, one can well wonder how much other damage accompanied the development of the forces leading to the running away, and to what extent the boys' personality structure would lead to even graver incidents.

The prognosis for these boys seems linked to the reaction of their parents. Granted that the effects of home changes will have limited influence on an adolescent close to adulthood, yet one gets the impression that most of the fathers and mothers conducted themselves in a way to give hope that in most instances they would be as helpful to their sons as possible.

It would be interesting to have as full a report on the way school people reacted to these same boys upon their return to school. Did they also take a forgiving attitude? Did they see the truancy as a sign they had a job to do? At this point the police records are silent.

SUMMARY

An investigation was made of five hundred and seventy-five boys reported missing from their homes. There seemed to have been an interplay of several home and school factors leading up to the trancies. Outstanding was the frequency with which search for adventure was a key factor. In most instances the aftermath found parents trying to better their relationships with their sons.

REFERENCES

- 1) Clairette P. Armstrong, "A Psychoneurotic Reaction of Delinquent Boys and Girls," *Journal of Abnormal and Social Psychology*, 32: 329-342, 1937.

- 2) Clairette P. Armstrong, *660 Runaway Boys*. Boston, Gorham Press, 1932.
- 3) Clairette P. Armstrong, *Why Boys Desert Their Homes*. Boston, Badger, 1932.
- 4) Herman M. Jahr, "Running Away," *Hygeia*, 18: 145-148, 156-157, 1940.
- 5) Owen R. Lovejoy, "Uncle Sam's Runaway Boys," *Survey*, 69: 99-101, 1933.
- 6) Towne Nylander, "Wandering Youth," *Sociology and Social Research*, 17: 560-568, 1932-33.
- 7) George E. Outland, "Relationship between School Drop-Outs and Boy Transiency," *Educational Administration and Supervision*, 23: 612-621, 1937.
- 8) Hugo Staub, "A Runaway from Home," *Psychoanalytical Quarterly*, 12: 1-23, 1943.
- 9) T. Earl Sullenger, *Social Determinants in Juvenile Delinquency*. New York, John Wiley & Sons, 1936.

A NOTE ON THE USE OF A "DON'T KNOW" ALTERNATIVE IN MULTIPLE CHOICE TESTS

MORTON P. FRIEDMAN and EDWIN A. FLEISHMAN¹

Skill Components Research Laboratory, Air Force Personnel and Training
Research Center, Lackland Air Force Base

A number of previous investigations (1-4) have shown that increasing the number of alternative responses in tests with multiple choice items increases the reliability of the total test. This has been shown true for two to seven alternatives and for vocabulary and arithmetic tests as well as an attitude inventory. The assumption in these previous studies was that the added alternatives are parallel in form and content to the original alternatives, as well as equally attractive as the original distractors, and that the increase in alternatives not only reduces the error variance due to guessing but also increases the number of judgments the examinee is required to make. In fact, it was found that the increase in reliability is roughly proportional to the increase predicted in the Spearman-Brown formula, letting the ratio of alternatives be substituted for the ratio of items.

There are, however, certain kinds of aptitude tests which call for a simple type of psychophysical judgment or discrimination, where the response is either "same or different," "yes or no," "higher or lower," etc. In this type of test it is often not feasible, and may not be technically possible, to add parallel alternatives to each item. Moreover, the reliability of such two choice judgments is notoriously low due to the amount of guessing and consequent chance variance contributing to score.

The present study was designed to test the hypothesis that reliability of tests of this type can be increased by simply providing

¹ This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project 7700. The opinions or conclusions contained in this report are those of the authors. They do not necessarily represent the views or endorsement of the Department of the Air Force. Millard M. Roberts provided valuable assistance in the conduct of the study.

the examinees with an additional "don't know" category for each item. This is really a special case of the two-choice situation in which the examinee is instructed to omit items he is not sure of and penalized for guessing through a scoring formula, although requiring the examinee to mark in a "don't know" choice presents a more structured situation. Moreover, providing an actual "don't know" choice on the answer sheet is especially important in the case of tests like the Seashore Measures of Musical Talents where items are presented aurally and the examinee must keep up with the items in turn under a fixed and moderately speeded space. In cases such as this, simply instructing examinees to omit items "they are not sure of" is out of the question since omitting items would frequently result in the examinee's "losing his place" on the answer sheet as he proceeds from item to item. In any case, the effects on reliability of these types of instructions and procedures have not been sufficiently charted.

It may be noted that a methodological advantage of using a test requiring mainly sensory discrimination is that the assumption that alternative responses are equally attractive is more likely to be satisfied.

PROCEDURE

The test used was a variation of the Rhythm Discrimination subtest of the Seashore Measures of Musical Talents (5). The present test is an auditory measure composed of fifty pairs of rhythmic patterns which increase in difficulty in groups of ten items.² The subject's task is to indicate if the patterns of beats within each pair have the "same" or a "different" rhythm. As used in the present study all instructions, examples, and test items are recorded on high-fidelity magnetic tapes.

The subjects were five hundred and ninety-four basic trainee airmen tested at Lackland Air Force Base. The subjects were randomly assigned to one of two groups. Group I received a two-choice IBM answer sheet, with only S (same) or D (different) slots next to each item number; Group II received a three-choice IBM answer sheet, where the extra choice was headed by a question mark. Instructions to Group I emphasized they were to answer

² The present test includes items in the Seashore Rhythm Test Series B as well as the more difficult items in the Series A test. Both the original Series A and B tests contain only thirty items each.

every item under either the S or D, even if they had to guess on certain items. Group II was instructed that they would be penalized for guessing and if they were not sure of their answer they were to mark under the question mark. This group also was instructed not to leave any item blank. Each group then received the identical fifty-item test as recorded on the tape.

RESULTS AND DISCUSSION

We are mainly concerned with the effects of the additional question mark response choice on reliability of the "Number Rights" score. Table I presents the odd-even reliabilities attained in each group. It can be seen that a slightly higher reliability was achieved for Group II, which was provided with the question mark category. The difference in coefficients is significant at the one per cent level. Table I also compares the reliability of the rights score alone versus the reliability of the rights minus wrongs score for this same test. It can be seen that use of the scoring formula results in no advantage in reliability over a simple rights score in this situation. In fact the R score has a corrected split-half reliability of 0.81 compared to a reliability of 0.80 for the R-W score.

It might also be noted that the correlation between total rights and total wrongs for Group II is -0.81 . This is high enough to

TABLE I—COMPARISON OF MEANS, STANDARD DEVIATIONS, AND RELIABILITIES OF SCORES OBTAINED FROM EACH GROUP ON THE TOTAL TEST

Score	No Question Mark Provided Group I (N = 288)				Question Mark Provided Group II (N = 305)			
	Mean	Standard deviation	Reliability* (uncorrected)	Reliability† (corrected)	Mean	Standard deviation	Reliability* (uncorrected)	Reliability† (corrected)
Rights	38.3	5.4	0.63	0.77	36.4	7.5	0.68	0.81
Rights minus wrongs					25.9	11.3	0.66	0.80

* Correlation between odd- and even-numbered items.

† Corrected for full length by the Spearman-Brown formula.

warrant the assumption that both scores measure the same abilities.

The distribution of question mark responses for Group II was markedly skewed; that is, a large number of subjects do not use the question mark at all. Of those subjects above the mean on total rights, the mean number of question marks used was 1.8 with only fifty-three per cent of the subjects using the question mark at all. On the other hand, for those subjects scoring below the mean on number of rights, the mean number of question marks used was 3.1 with seventy-eight per cent of these subjects using the question mark. Thus, it is obvious that those subjects who have less ability on the trait in question, and who would have contributed most of the error variance through guessing on a two-choice test, have made the most use of the question mark, leaving a more reliable rights score.

A further check was made on the effect of test difficulty on the utility of the question mark response. The hypothesis here was that the use of question mark response results in a greater improvement in reliability as the difficulty of a test increases. To test this hypothesis a parallel analysis was carried out on the scores achieved on the last thirty items of the test, which were the most difficult items. Means, standard deviations, and odd-even reliabilities are presented in Table II. Table II indicates that the differences

TABLE II—COMPARISON OF MEANS, STANDARD DEVIATIONS, AND RELIABILITIES OF SCORES OBTAINED FROM EACH GROUP ON THE MORE DIFFICULT THIRTY ITEMS

Score	No Question Mark Provided Group I (N = 288)				Question Mark Provided Group II (N = 305)			
	Mean	Standard deviation	Reliability* (uncorrected)	Reliability† (corrected)	Mean	Standard deviation	Reliability* (uncorrected)	Reliability† (corrected)
Rights	21.8	3.6	0.53	0.70	20.5	4.5	0.61	0.76
Rights minus wrongs					13.4	7.7	0.56	0.71

* Correlation between odd- and even-numbered items.

† Corrected for full length by the Spearman-Brown formula.

between Groups I and II in obtained reliabilities is much more pronounced when only this more difficult part of the test is evaluated. An increase of more than 0.06 is achieved when the question mark choice is provided. It is also shown that the reliability of the rights only score is again superior to the reliability of the R-W score, and that this superiority is greater for the more difficult test. All of these differences are statistically significant beyond the one per cent level.

It is to be noted that the fifty-item test utilized above had a difficulty (P) of 0.52 (estimated from R-W where question mark is allowed) which is in the general range likely to be used for maximum discrimination. The more difficult thirty-item test had a difficulty of 0.45. Our conclusions regarding the benefits to reliability of adding a question mark are consequently limited to difficulty levels in the "average to moderately difficult" range. It appears obvious, however, that for tests which sample a narrow range of difficulty, either at the "too easy" or "too difficult" end, the use of the question mark has doubtful utility. Examinees would avoid the question mark on the easy test and the number of question mark responses might be so great on the difficult test as to make the total correct score useless. Moreover, it is not known if similar results would be obtained with different types of tests or even other tests involving psychophysical types of judgment. It is likely, however, that our findings would hold for similar tests of the latter kind. It should also be stressed that the present study has little bearing on other studies on the utility of question mark responses with reference to personality measurement, where the question mark response represents a neutral category and there is often interest in keying this response as a separate measure of "caution" or some similar trait.

SUMMARY

The effects of adding a "don't know" response choice to an aural discrimination test consisting of two-choice items was examined. The results indicate a significant increase in reliability of the total rights score as a function of adding the "don't know" alternative. Moreover, this increase in reliability was most pronounced for the most difficult part of the test. It was also found that, in the situation where a "don't know" response was allowed,

the rights score alone was either equal or superior in reliability to a rights minus wrongs score.

The conclusions, although limited to the range of difficulty and kinds of items employed, have special implications where one wishes to raise the reliability of such a test, but its inclusion as part of a larger battery with fixed time limits precludes the inclusion of additional items.

REFERENCES

- 1) H. R. Denny and H. H. Remmers, "Reliability of Multiple-Choice Measuring Instruments as a Function of the Spearman-Brown Prophecy Formula, II," *Journal of Educational Psychology*, 31: 699-704, 1940.
- 2) L. B. Plumlee, "The Effect of Difficulty and Chance Success on Item-Test Correlation and on Test Reliability," *Psychometrika*, 17: 69-86, 1952.
- 3) H. H. Remmers and E. Ewart, "Reliability of Multiple-Choice Measuring Instruments as a Function of the Spearman-Brown Prophecy Formula, III," *Journal of Educational Psychology*, 32: 61-66, 1941.
- 4) H. H. Remmers and J. M. House, "Reliability of Multiple-Choice Measuring Instruments as a Function of the Spearman-Brown Prophecy Formula, IV," *Journal of Educational Psychology*, 32: 372-376, 1941.
- 5) C. E. Seashore, D. Lewis and J. G. Saetveit, *Manual of Instructions and Interpretations for the Seashore Measures of Musical Talents* (1939 revision). Camden, New Jersey, Educational Department, RCA Victor Division, Radio Corporation of America, 1942.

PROBLEM BEHAVIOR IN SCHOOL

MERRILL T. EATON

Indiana University

LOUIS A. D'AMICO

Xavier University (New Orleans)

and

BEEAMAN N. PHILLIPS

The University of Texas

Juvenile delinquents have often shown milder forms of undesirable or maladjustive behavior prior to their delinquent behavior (8, 9). The youth who at the age of sixteen is in court for stealing a car frequently is the youth who at the age of ten stole things, lied, and engaged in other misbehavior at home, at school, and in the community. Therefore, the school which strives to control and prevent problem behavior, and which fosters good adjustment in all youth while they are still in school, will make a substantial contribution toward reducing the amount of delinquency in the community.

There have been only a relatively small number of research studies on problem behavior in school. One of the best is the pioneer study by Olson (10). Other studies have been conducted by Epstein (3), Sparks (12), Stauffer (13), Hildreth (4), Laycock (6), Wickman (14), Yourman (15), McClure (7), Boynton and McGaw (1), Dickson (2), and Peck (11).

A review of these studies indicates that there are many gaps in our knowledge of various aspects of such problem behavior. On the basis of available information we cannot answer such questions as: Why are pupils problems in some classes and not in others? How does problem behavior differ from grade level to grade level? What types of problem behavior occur most frequently among boys and girls in different age groups? and a host of other questions.

Problem behavior occurring in school and out of school was the subject of the investigation reported here. The specific purposes of this study were to answer the following questions. Does the

frequency of problem behavior increase from grades one to six, to grades seven to nine, to grades ten to twelve? Do different types of problem behavior predominate in these grades? Are boys involved in more problem behavior than girls at all grade levels? Is there a difference in the types of problem behavior occurring most frequently among boys and girls in these grades?

PROCEDURES

Thirty teachers were interviewed during the summer of 1954, all of whom had taught during the preceding school year. Some were from the same school system, but no two teachers were from the same school. Twelve of the teachers taught in grades one to six, ten taught in grades seven to nine, and the remainder taught in grades ten to twelve. During the school year they taught approximately two thousand different pupils. These pupils were distributed as follows: four hundred in grades one to six, six hundred in grades seven to nine, and one thousand in grades ten to twelve.

A special form was developed and used in the interview. This form provided information on the frequency of such in-school problem behavior as interrupting, carelessness, inattention, "smartness," or silliness, impertinence and impudence, unexcused absences, cheating, fighting, smoking, lying, stealing, swearing, and obscene language. Information also was provided on the frequency of such out-of-school problem behavior as drinking, illicit sex activities, and vandalism. In addition, the number of pupils expelled from school, sent to state schools, or picked up by the police for various reasons was obtained.

The teacher determined what behavior should be classified as "problem behavior." For example, when the question, "Did you have any pupils who were problems because of swearing?" was asked, it was the responsibility of the teacher to decide how many of her pupils had been problems during the year because of the behavior named. Admittedly, there are some obvious limitations to this procedure. Teachers differ in the way they look at pupil behavior and in what they would consider to be "problem behavior." However, this is the way such behavior usually is identified, so for the purpose of this study the teacher's judgment in this matter was considered sufficient.

All problem behavior was tabulated according to type, grade

level, and the sex of the pupils. These frequencies were then converted into percentages, or the frequency of occurrence among every one hundred pupils.

Since the data were assembled in a subjective manner no attempt was made to test the statistical significance of the differences. For this reason the findings should be thought of as tentative or suggestive.

RESULTS

Does problem behavior increase with grade level? One of the purposes of this study was to determine whether or not there is an increase in either in-school or out-of-school problem behavior from grades one to six to grades ten to twelve. The frequency of occurrence of all types of problem behavior per one hundred pupils in these grades is shown in Table I.

It is evident that there was an increase in problem behavior from grades one to six to grades ten to twelve. The increase in in-school problem behavior was greatest between grades one to six and seven to nine; while the increase in out-of-school problem behavior was greatest between grades seven to nine and ten to twelve.

TABLE I. FREQUENCY OF OCCURRENCE OF ALL TYPES OF PROBLEM BEHAVIOR PER ONE HUNDRED PUPILS

Place	Grades		
	1-6	7-9	10-12
In-school	31	40	43
Out-of-school	1	2	12
Both	32	42	55

TABLE II. FREQUENCY OF OCCURRENCE OF ALL TYPES OF PROBLEM BEHAVIOR PER ONE HUNDRED PUPILS

	Grades		
	1-6	7-9	10-12
Boys	40	50	91
Girls	23	35	30

Do boys engage in more problem behavior than girls? Another purpose of this study was to determine whether or not boys are involved in more problem behavior than girls. The data in Table II show the frequency of occurrence of all types of problem behavior per one hundred boys and girls in these grades.

It is obvious from the data that boys were involved in more problem behavior than girls at all three grade levels. The differences, however, were greatest at the senior high school level and least at the junior high school level.

Does problem behavior vary from grade level to grade level? A third purpose of this study was to determine whether or not there are variations in the types of problem behavior occurring most frequently at each grade level. The frequency of occurrence of seventeen major types of problem behavior was computed for grades one to six, seven to nine, and ten to twelve and are shown in Table III.

TABLE III. FREQUENCY OF OCCURRENCE OF SEVENTEEN MAJOR TYPES OF PROBLEM BEHAVIOR PER ONE HUNDRED PUPILS

Activity	Grades		
	1-6	7-9	10-12
Carelessness in work	9.1	8.3	3.3
Inattention	4.6	3.8	2.1
Cheating	2.2	1.9	4.5
Smoking	1.2	1.9	4.8
Restlessness	1.2	6.0	0
Interrupting	1.2	3.8	2.1
Unexcused absences	1.6	2.4	5.0
Lying	1.2	0.8	0
Stealing	1.2	0.9	5.0
Smartness	1.6	3.6	1.0
Whispering, notewriting	1.6	5.1	0
Swearing	0.6	0.6	7.6
Obscene language	1.2	0.3	1.1
Disorderliness	0.4	1.7	0.3
Drinking	0	0.5	7.4
Illicit sex activities	0.4	0.9	1.4
Vandalism	1.2	0.3	0.1

The data indicate that most types of problem behavior varied in frequency from grade level to grade level. Such problem behavior as carelessness in work, inattention, lying, and vandalism decreased in frequency from grades one to six to ten to twelve. Other problem behavior such as smoking, unexcused absences, stealing, swearing, drinking, and illicit sex activities increased in frequency from grades one to six to ten to twelve. Still other problem behavior such as restlessness, interrupting, smartness, whispering and notewriting, and disorderliness occurred most frequently in grades seven to nine.

Is sex related to the frequency of various types of problem behavior at different grade levels? The last purpose of this study was to determine if there are differences in the types of problem behavior occurring among boys and girls at each grade level. Differences in frequency for the seven most common types of problem behavior are shown in Table IV for each grade level in rank order for boys.

The data below clearly reveal not only that there was more problem behavior among boys, but also that there were more boys involved in almost every type of problem behavior. The only important exceptions to this were smoking and illicit sex activities (not shown). Frequently occurring problem behavior which was more or less characteristic of both boys and girls included carelessness in work, inattention, restlessness, and interrupting.

Data collected from the teachers also revealed that twenty-five pupils were picked up by the police, twenty-two were sent to a

TABLE IV. DIFFERENCES IN FREQUENCY FOR SEVEN MOST COMMON TYPES OF PROBLEM BEHAVIOR IN RANK ORDER FOR BOYS

Grades 1-6		Boys	Girls	Grades 7-9		Boys	Girls	Grades 10-12		Boys	Girls
Carelessness in work		9.5	8.4	Carelessness in work		9.3	7.1	Swearing		15.6	1.5
Inattention		5.1	3.7	Restlessness		6.1	5.8	Unexcused absences		9.1	1.9
Smartness		2.2	0.5	Whispering, notes		4.6	2.9	Stealing		9.1	2.1
Fighting		2.2	0.5	Unexcused absences		2.9	1.7	Cheating		7.4	2.3
Smoking		1.9	0.0	Cheating		2.7	0.9	Smoking		3.2	6.0
Stealing		1.9	0.0	Smoking		2.2	1.4	Interrupting		3.2	1.3
Interrupting		1.9	0.0	Stealing		1.4	0.1	Inattention		3.0	1.5

state home, and nineteen were expelled from school. Some of the reasons given for this were stealing, gambling, arson, running away from home, burglary, using narcotics, illicit sex activities, and violations of driving laws.

DISCUSSION

On further analysis and interpretation it becomes apparent that the majority of the types of problem behavior reported by the teachers can be classified either as violations of classroom work and behavior requirements or as violations of moral standards. When all the problem behavior which was reported was classified in this way, it was found that violations of classroom work and behavior requirements made up sixty per cent of the problem behavior occurring in school. When the analysis was made by grade levels, violations of classroom work and behavior requirements constituted seventy-three per cent of the problem behavior occurring in grades one to six, seventy-seven per cent of the problem behavior occurring in grades seven to nine, but only twenty-eight per cent of the problem behavior occurring in grades ten to twelve.

In view of the large amount of problem behavior that is related to discipline and motivation, it is reasonable to assume that teachers need more training and guidance in these two important areas of teaching. For teachers already in service, this might be accomplished through workshops explaining causes and ways of alleviating motivational and discipline problems. For prospective teachers, more supervised contact with children, more instruction in teaching techniques and psychological principles and applications, and more individualized guidance and counseling might be helpful.

Perhaps the most significant aspects of this study are the findings which suggest that the types of problem behavior occurring in a classroom may be related to different phases of growth and development. For example, it was found that problem behavior among girls reaches a peak in grades seven to nine, and that problem behavior among boys reaches a peak in grades ten to twelve. It might be just a coincidence that these grade periods correspond roughly with the beginning of adolescence of boys and girls, but on the other hand, there might be a relationship between the "storms and stresses" of adolescence and increases in certain types of problem

behavior in school. Similarly, in view of the confining nature of most classroom situations it is quite possible that increases in restlessness, inattentiveness, and generally disorderly behavior are related to the period of rapid physical growth and development of boys and girls.

Further research is needed not only to substantiate the results of this study but also to investigate other aspects of problem behavior in school. For example, is there actually a change from grade to grade in the behavior of pupils (as found in this study), or do these apparent changes result from the fact that behavior which is annoying at one grade level is not annoying at another. As another example, there has been little research on the nature of problem behavior. It is not entirely clear as to how much the meaning of the term "problem behavior" varies from teacher to teacher. Nor is it clear as to whether or not differences in the amount of problem behavior reported by different teachers represent actual or conceptual differences. And as a third example, there hasn't been much research on the causes of problem behavior in school. There appear to be differences in the types and the amounts of problem behavior occurring at different grade levels, but there is no clear-cut evidence as to what causes these differences. Perhaps they are associated with such factors as teaching techniques, growth and development, the teacher's personality, and pupil-teacher and pupil-pupil relationships.

However, the literature is clear on one thing. The degree to which we are able to reduce and control school-related problem behavior depends on the degree to which we understand it. It is evident, therefore, that we must increase our efforts to understand both the nature and the causes of problem behavior.

REFERENCES

- 1) P. L. Boynton and B. H. McGaw, "The Characteristics of Problem Children," *Journal of Juvenile Research*, 18: 215-22, 1934.
- 2) V. E. Dickson, "Behavior Difficulties That Baffle Teachers," *Journal of Juvenile Research*, 16: 93-101, 1932.
- 3) L. J. Epstein, "An Analysis of Teachers' Judgments of Problem Children," *Pedagogical Seminary and Journal of Genetic Psychology*, 59: 101-107, 1941.
- 4) G. Hildreth, "A Survey of Problem Pupils," *Journal of Educational Research*, 18: 1-19, 1928.
- 5) A. Hill and others, "Schools Face the Delinquency Problem," *Na-*

tional Association of Secondary School Principals Bulletin, 37: 181-221, December, 1953.

6) S. R. Laycock, "Teachers' Reactions to Maladjustments of School Children," *British Journal of Educational Psychology*, 4: 11-29, 1934.

7) W. E. McClure, "Characteristics of Problem Children Based on Judgments of Teachers," *Journal of Juvenile Research*, 13: 124-40, 1929.

8) National Society For the Study of Education, *Juvenile Delinquency and the Schools*, Forty-Seventh Yearbook, Part I, 280 pp. Chicago, University of Chicago Press, 1948.

9) New York State Youth Commission, *Reducing Juvenile Delinquency*, 34 pp. Albany, New York, The Commission, 1952.

10) Willard C. Olson, *Child Development*. New York, D. C. Heath, 1949.

11) L. Peck, "Teachers' Reports of the Problems of Unadjusted School Children," *Journal of Educational Psychology*, 26: 123-38, 1935.

12) J. N. Sparks, "Teacher Attitudes Toward the Behavior Problems of Children," *Journal of Educational Psychology*, 43: 284-91, 1952.

13) G. A. W. Stauffer, "Behavior Problems of Children as Viewed by Teachers and Mental Hygienists: A Study of Present Attitudes as Compared With Those Presented by E. K. Wickman," *Mental Hygiene*, 36: 271-85, 1952.

14) E. K. Wickman, *Children's Behavior and Teachers' Attitudes*, 247 pp. New York, The Commonwealth Fund, Division of Publications, 1928.

15) Julius Yourman, "Children Identified by Their Teachers as Problems," *Journal of Educational Sociology*, 5: 334-43, 1932.

EFFECT OF ANGULAR ALIGNMENT UPON READABILITY OF PRINT¹

MILES A. TINKER

University of Minnesota

When printed copy is arranged to have the lines run horizontally left to right so that they are perpendicular to the median plane of the body, eye movements in reading are relatively uncomplicated. The eyes merely move horizontally from one fixation pause to the next along a line of print. These saccadic eye movements are probably the most practiced oculomotor adjustments made by any person who does much reading. The interior and exterior recti of the extrinsic eye muscles do most of the work during such movements. When the printed copy is rotated less than 90 degrees away from this horizontal position, the oculomotor adjustments in reading become more complicated. The eyes have to move obliquely from one fixation to another along the line of print. This involves more extrinsic eye muscles and is a more intricate and more difficult type of oculomotor adjustment than moving right and left horizontally. Also with the oblique alignment of printed lines, the word forms are in an unfamiliar orientation. If the printed lines run vertically, either up or down, the eye movements are not as difficult to execute as in the oblique alignment but more difficult than in the horizontal. In addition the unfamiliar orientation of the word forms in the vertical alignment may well hinder rapid reading.

In many reading situations, the alignment of printed lines frequently departs to some degree from the horizontal position. This occurs with both children and adults, particularly when a book or magazine is lying flat on a table or desk and the reader is taking notes. And many backbone titles of books and magazines are printed in a vertical alignment. Some of the latter have to be read upward, others downward. Any complication which interferes with smooth eye movements or visibility and hence with rapid perception is likely to slow up speed of reading.

The specific problem in this study is to investigate changes that

¹ This research was supported by a grant from the Graduate School, University of Minnesota.

occur in speed of reading and in visibility of words when the alignment of printed lines departs from the horizontal. In this preliminary study, the effects of rather marked departure from the horizontal alignment will be investigated.

MATERIALS AND PROCEDURE

The reading material consisted of eighteen paragraphs in each of Forms A and B of the Chapman-Cook Speed of Reading Test. Typography was identical in the two forms (ten-point Old Style set solid in a nineteen-pica line width on eggshell paper stock). A specially constructed rotating reading stand was constructed. The circular surface of the stand was forty-five degrees to the table top and so arranged that the printed copy could be rotated and held at any degree from the horizontal to the vertical alignment. The printed material was mounted in two columns on white cardboard which was attached to the rotating surface so that the center of the printed area was exactly over the center of the rotating stand.

For the visibility measurements, sixteen five-letter words that occur in both Form A and Form B were cut from the test and mounted on white cards, one word per card. Each card was attached to the rotating table so that the center of a word was exactly over the center of the table. Visibility was measured by the Luckiesh-Moss Visibility Meter.

Five test groups of sixty college sophomores each were used. The experiment was conducted in a special laboratory with twenty foot-candles of well-diffused indirect light. Each subject was tested individually. On arrival at the laboratory, a subject was told the purpose of the experiment and given the visibility measurements followed by the speed of reading test. The subject's head was maintained in approximately a constant position by means of a headrest. There were practice exercises for the reading test and for the visibility measurements. In Test Group I, the control group, the subjects read Form A and then B, both at the horizontal alignment, i.e., at 0° . Time was recorded for reading the last fifteen of the eighteen paragraphs. Each paragraph contained thirty words and the subject responded by speaking aloud the one word that spoiled the meaning of the paragraph. After four practice words, two sets of six words each were read through the visibility meter at the 0° position. Data from Group I revealed

degree of equivalence in responding to test Forms A and B and to the two sets of words. In all the experimental groups, Group II through V, Form A was read at 0° and visibility measures for the first six words (after practice set) were taken at 0° . In Group II, Form B and the last six words for the visibility measurements were aligned at 45°C (rotated clockwise). This required reading on a slant from upper left to lower right. In Group III, the alignment was at 90°C (vertical) so that reading was done from top to bottom. In Group IV, the alignment was -45°CC (counterclockwise) so that reading was on a slant from lower left to upper right. And in Group V the alignment was -90°CC so that material was read in the vertical from bottom to top.

RESULTS AND DISCUSSION

The basic data for the speed of reading measurements are given in Table I. The mean scores in column 4 represent the time in

TABLE I—EFFECT OF ANGULAR ALIGNMENT OF PRINT UPON SPEED OF READING

Test Group	Test Form	Alignment of Print	Mean in Seconds	SE _M	Difference Between Corrected Means* in		r	D SE _D
					Seconds	Per cent		
I	A	0	85.11	3.71	0.00	0.0	0.94	—
	B	0	88.73	3.59				
II	A	0	79.52	2.80	-40.90	-51.4	0.78	15.81
	B	45°C	116.78	4.06				
III	A	0	80.05	2.34	-162.85	-203.4	0.26	13.70
	B	90°C	239.27	12.27				
IV	A	0	82.28	3.12	-43.72	-53.1	0.75	13.79
	B	-45°CC	122.37	4.74				
V	A	0	79.18	2.69	-163.70	-206.7	0.25	17.35
	B	-90°CC	239.25	9.74				

The mean score is the average number of seconds taken to read 15 paragraphs of 30 words each in the Chapman-Cook Speed of Reading Test. In each test group $N = 60$ college students, 300 in all.

* The difference between the means in each of the groups II, III, IV, V is "corrected" by the amount of the difference between the means in the control group (-3.63 sec.). C refers to clockwise rotation; CC to counterclockwise.

seconds for reading fifteen paragraphs of thirty words each. When the text was rotated 45°C (clockwise), speed of reading was retarded by 51.4 per cent. When rotated counterclockwise by the same amount (-45°CC), the retardation was about the same, i.e., 53.1 per cent. In reading vertically downward (90°C), the retardation was 203.4 per cent; while reading vertically upward (-90°CC) it was 206.7 per cent. So we find that reading material which deviates markedly from the horizontal alignment retards speed of reading by a large amount.

Data for the visibility measurements are given in Table II. The trend of the results is similar to those for speed of reading but not as pronounced. Rotating words 45° in either direction reduced visibility by eleven to twelve per cent, and rotating the words to a vertical alignment (90°), decreased visibility by about twenty-one per cent. All differences are highly significant in both tables.

TABLE II—EFFECT OF ANGULAR ALIGNMENT OF WORDS UPON VISIBILITY OF PRINT

Group	Alignment of Print	Mean Visibility Score	SEM	Difference Between Corrected Means*		r	D SED
				Score	Per cent		
I	0	4.04	0.13	0.00	0.0	0.95	—
	0	4.03	0.13				
II	0	4.39	0.16	-0.48	-11.1	0.92	7.99
	45°C	3.90	0.14				
III	0	4.00	0.13	-0.84	-20.9	0.73	9.07
	90°C	3.15	0.11				
IV	0	4.12	0.12	-0.50	-12.2	0.88	8.28
	-45°CC	3.61	0.11				
V	0	4.11	0.13	-0.85	-20.7	0.85	12.05
	-90°CC	3.25	0.11				

The mean visibility score is the average for perceiving six words by use of the Luckiesh-Moss Visibility Meter. In each test group $N = 60$ college students, 300 in all.

* The difference between the means in each of the groups II, III, IV, and V is "corrected" by the amount of the difference between the means in the control group ($+0.0098$).

The results presented above demonstrate that readability of print is adversely affected by a significant amount when the alignment of the printed lines departs markedly from the horizontal. Speed of reading is affected relatively much more than visibility of the printed words. In addition to reduced visibility (ease of perceiving words), other hampering factors are likely involved in slowing up speed of reading: (a) The interfixation eye movements require complicated oculomotor adjustments in the 45° oblique alignment. It is also generally recognized that precisely adjusted vertical eye movements are more difficult to execute than horizontal movements. (b) In the horizontal (0°) alignment, fast reading depends to a considerable degree upon effective use of peripheral vision to the right of a fixation pause. Then phrasing or reading by thought units is facilitated. Apparently effective use of peripheral vision is inhibited when the print is oblique or vertical. The reader is reduced pretty much to word-by-word reading which is slow. (c) In ordinary reading, a majority of words are quickly recognized by their configuration or total word form. Although the over-all shape of word forms is the same whether in horizontal, oblique or vertical alignment, their orientation in the oblique or vertical alignment makes it difficult to recognize them at a glance.

It would seem, therefore, that the reduced speed of reading material which deviates from a horizontal alignment may be due, in part at least, to the complicated eye movements involved, to reduced visibility, to inability to make effective use of peripheral vision, and to reduction of word-form clues. The contribution of each of these hindering factors is not clear from the results of this experiment. Certainly, there must be other factors than visibility operating as seen by comparing the results in the two tables. The percentage reduction in visibility is much less than retardation in speed of reading.

It may be emphasized that direction of rotating the copy, clockwise or counterclockwise, makes little difference. It is as easy to read down as up vertically, and from upper left to lower right as from lower left to upper right obliquely. Apparently publishers may print backbone titles upward or downward. One arrangement is as good (or as bad) as the other.

Further investigation is needed to determine precisely how much print may be rotated away from the horizontal alignment before ease and speed of reading is adversely affected by a significant

amount. The results in this experiment are striking enough so that it is advisable for readers, including school children, to maintain the alignment of print approximately in the horizontal position while reading.

SUMMARY

(1) The purpose of this study is to determine the changes in speed of reading and visibility of words when the angular alignment of printed lines departs from the horizontal.

(2) The subjects were three hundred college students divided into five subgroups of sixty each.

(3) Each subject was timed for reading fifteen thirty-word paragraphs in the horizontal alignment and another fifteen in one of the following: horizontal, 45° clockwise, vertical reading downward, -45° counterclockwise, vertical reading upward. Visibility of words was measured in each of the same alignments.

(4) The 45° alignment retarded speed of reading by about fifty per cent, and the vertical alignment by about two hundred per cent.

(5) The 45° alignment decreased visibility by eleven to twelve per cent and the vertical alignment by about twenty-one per cent.

(6) It is suggested that, in addition to reduced visibility, the retardation in speed of reading when the alignment deviated from the horizontal was likely due partly to unfamiliar orientation of word forms, inability to use peripheral vision effectively, and increased complexity of oculomotor requirements.

(7) Since print can be read as well upward as downward, one alignment may be used as well as another for backbone titles on books and magazines.

(8) It is suggested that printed material be held so that there is no marked deviation from the horizontal when reading. The safe margin of such deviation has not been determined in this experiment.

TEACHER-PUPIL RAPPORT AND STUDENT TEACHER CHARACTERISTICS

GALEN W. DODGE AND DONALD O. CLIFTON

University of Nebraska

If one were willing to assume that the most effective teachers are well thought of by their students, or that it would be worthwhile to know if there is anything unique in the personality of a teacher who develops rapport with his students, then this article is of interest.

The purpose of this study was to determine the relationship between the degree of rapport developed by student teachers with their pupils and certain other characteristics of these student teachers. The degree to which rapport was established was operationally defined by using the score obtained when the pupils listed the three teachers they liked best, learned the most from, and obtained help from the easiest. This score was related to the following student teacher characteristics: peer ratings, supervisors' ratings, ratings as prospective teachers when they were in a freshman orientation course, cumulative grade averages, and scores on the American Council on Education Psychological Examination (3).

Statements of the need for such a study and reviews of the research to date have been previously published by A. S. Barr (2).

PROCEDURES

The first task was to devise a questionnaire which would measure the degree to which teacher-pupil rapport was established, so that this result could be correlated with other student teacher characteristics.

The development, administration and scoring of the questionnaire. The first step in the development of the questionnaire was the selection of positive statements which would indicate the degree of rapport between the student teachers and the pupils. The next step was to provide directions which would permit the pupil to express himself as honestly as possible.

The three statements selected were as follows:

- (a) Of the teachers you now have, choose three you like the most.

(b) Of the teachers you now have, choose three from whom you learn the most.

(c) Of the teachers you now have, choose three from whom it is easiest to obtain help.

The directions were as follows:

The purpose of these ratings is for research. We want to find out what good teachers are like. Since the results will depend on how well you give your answers, your coöperation will be greatly appreciated. Your teacher will not see your rating or the results of the ratings. It is not necessary for you to sign your papers and the ratings will not affect you or your teachers. First, turn to pages 2 and 3 and place an X beside the name of every teacher that you have now. Second, from the teachers which you have checked, answer the questions below. List your teachers in order of preference: first choice, first; second choice, second; and third choice, third. Use only the names that you have checked on pages 2 and 3. You may use the same names for different questions if you wish. You should be finished in five minutes, so work quickly. When you are finished, fold your paper.

The pupils were asked to read the directions while they were read aloud by the administrator. As the directions indicate, a list of all the student teachers was given to each student from which he was to select those he had at that time. This was done to call his attention to all the teachers he had, to reduce such errors as listing teachers of previous semesters, and to help students who were unsure of the spelling of their teachers' names. This list also gave a check on the total possible number of ratings which a teacher might receive. In addition to the assurance that giving frank opinions would not affect the pupils adversely, the questionnaires were collected by persons who were not associated with the high school staff.

The questionnaire was administered at the end of both the first and second semesters. The scores on both administrations were pooled and treated statistically as one population.

The degree to which each student teacher developed rapport was represented by a single total score called the Teacher-Pupil Rapport score (TPR). The TPR was obtained by summing for each student teacher the weighted responses—first choice, 3 points; second choice, 2 points; and third choice, 1 point—and then dividing by the total possible number of pupils by whom the teacher could have been chosen. The possible number of choices that a student teacher might receive was considered to be proportionate to the number in the class since almost all students took five

classes and each student teacher had only one class. This method of determining the score gave each student teacher the possibility of obtaining a maximum score of nine. The distribution of the TPR scores ranged from 0.00 to 8.08 with a mean of 3.22 and a standard deviation of 1.96.

The Experimental Situation

The availability of a favorable experimental situation gave impetus to this study. In the University High School at the University of Nebraska the teaching is performed almost entirely by student teachers under the direction of supervisors. This situation had the advantages of having a relatively large number of student teachers who could be rated and also a principal and faculty who were very receptive to this research. While the subjects were student teachers it seemed that there would be implications for teachers in the field since these student teachers have the primary responsibility for their classes throughout the semester. Each student teacher is graded by his supervisor and obtains credit from the University for his practice teaching.

The Sample

The sample of student teachers was obtained from the total of one hundred and twenty-five student teachers enrolled in either the first or second semester of the 1954-55 teacher training program at the University of Nebraska. From this sample it was possible to compute TPR scores for one hundred and twelve; fifty-eight females and fifty-four males. A t of 0.054 indicated no significant sex differences for the TPR scores. Three of the student teachers were omitted because they were also serving in staff positions at University High. The other ten were not included because they worked in programs of short duration, or, as in some of the music courses, they had little possibility for individual contact with their pupils. For nineteen students who taught both semesters only the first semester ratings were used.

The pupils who rated the student teachers consisted of the high school pupils enrolled in the Teachers College High School during the 1954-55 school year. This was a four-year high school at the time of this study composed of two hundred and twenty pupils, one hundred and eight girls and one hundred and twelve boys,

during the first semester; and two hundred and thirteen pupils, one hundred and six girls and one hundred and seven boys, during the second semester.

Techniques of Measurement

Several techniques were selected to measure student teacher characteristics which were hypothesized to be predictive of the pupils' ratings of the teacher. Each technique will be described briefly.

American Council on Education Psychological Examination (ACE). The ACE scores that the student teachers received when they were freshmen were obtained from the Bureau of Instructional Research at the University. This examination was designed to measure two factors of mental ability, the "L," linguistic, and the "Q," quantitative. As used at the University of Nebraska, the scores are reported on a 9-point scale with 1 being lowest and 9 the highest and a mean at 5.

Professors' estimates of future teaching ability. Students who come into the Teachers College with no previous college credit take the first course in Education at the University which is an orientation to the University and to the field of teaching. When this course was finished at the end of the first semester, the professors were asked to give a global rating of each student as a prospective teacher. As a basis for their estimates, the professors who taught this course had made it a point to become well acquainted with each of their students, and, in addition, had knowledge of the ACE scores, class participation, and grades on class tests. Each estimate was recorded using a 1-to-5 scale where 1 was low, 5 high, and the average was 3.

Supervisors' ratings. As described previously, each student teacher had a class which he conducted under the supervision of a regular staff member. At the conclusion of the semester a supervisor, using a 1 as low to a 5 as high, rated each student teacher on the following items: teaching methods, rapport with students, classroom management, interest in teaching, personal qualities, and an over-all rating considering all factors. These ratings were available for all the student teachers.

Student teaching grades. The student teachers were given a grade in student teaching by their supervisors. This grading system

consists of a 9-point rating scale where 9 is the highest grade. For this sample, the grades ranged from 6 to 9 with a mean of 7.34 where N was 111.

Cumulative grade average. The average of all the grades each of the student teachers had earned at the University or had transferred were obtained from the Office of Registration and Records. For all the student teachers this included at least six semesters of college work and in several instances seven. These grade averages ranged from 4.22 to 7.79 and the mean for one hundred and seven student teachers was 6.02.

Friendship Rating Scale. For thirty-three of the student teachers peer ratings were available which represented the mean ratings on nine characteristics from every other member of their social fraternity or sorority. This information was collected, using an instrument called the Friendship Rating Scale (1) which has the advantage of collecting evaluations of a person by forty to eighty people who see him daily and in a variety of real-life situations. The nine items on which every member of a house rated every other member on a five-point continuum are as follows:

- (a) He fits the following description: Very cheerful and has a sense of humor. Always tries to help someone else who may have troubles. Is honest and others trust him. Other people like him and try to be with him.
- (b) When he notices a shy person in a group who is not having any fun he tries to get this person into the group to have a good time.
- (c) He greets other people when he meets them.
- (d) He is a person with whom I would like to talk over my personal affairs and secrets.
- (e) One can depend upon him to keep a bargain.
- (f) He is generous in sharing his possessions with others.
- (g) He will be making a contribution to his community twenty years from now.
- (h) I consider him to be one of my best friends.
- (i) He talks about other people's good points rather than their bad ones.

This rating scale was administered during the spring semester of 1954 in nine sororities and eleven fraternities. The scores for the student teachers ranged from 2.75 to 4.28 with a mean of 3.48.

RESULTS

The results will be discussed with reference to the TPR questionnaire, per se, and to the relationships found between the TPR and other student teacher characteristics.

The TPR Questionnaire

The intercorrelations between the items of the TPR questionnaire where N was 111 were as follows: Between one and two, 0.80; between one and three, 0.89; and between two and three, 0.78. From this, one might assume that the pupils rate the teachers on all the items about the same depending on how well they like them, in general.

Correlations Between the TPR Scores and Characteristics of the Student Teachers

The Pearson Product-Moment method was used to determine the coefficients of correlation. Neither the ACE linguistic nor quantitative scores correlated significantly with the TPR scores. The coefficients of correlation where N was 77 were respectively 0.15 and 0.19. Although these correlations are probably depressed because the range of intelligence is limited within this group, it is nevertheless indicative of an almost negligible contribution to the prediction of teacher-pupil rapport.

The professors' estimates of future teaching ability correlated 0.59 with the TPR scores. For an N of only 31 this coefficient is significant at the one per cent level of confidence. This is a particularly interesting result when it is considered that the professors' estimates were made approximately three years previous to the pupils' ratings.

The coefficients of correlation between the supervisors' ratings of one hundred and eleven of the student teachers and the TPR scores are as follows: teaching results, 0.14; teaching methods, 0.09; rapport with students, 0.26; classroom management, 0.27; interest in teaching, 0.09; personal qualities, 0.31; and over-all, 0.30. Four of the items, rapport with students, classroom management, personal qualities, and the over-all are significant at the one per cent level of confidence.

The student teaching grades were correlated with the TPR scores for one hundred and eleven student teachers and a coeffi-

cient of 0.24 was obtained. This coefficient is statistically significant at the five per cent level of confidence.

The coefficient of correlation between the student teachers' cumulative grade average and the TPR scores was 0.25 where N was 107. This coefficient is significant at the one per cent level.

Each of the nine items and the total score of the Friendship Rating Scale were correlated with the TPR scores. The coefficients are as follows: (a) 0.42; (b) 0.56; (c) 0.54; (d) 0.31; (e) 0.34; (f) 0.34; (g) 0.56; (h) 0.32; (i) 0.44; and total score 0.50. Since N was 33, correlations for items (b), (c), (f), (i) and total are significant at the one per cent level and item (a) is significant at the five per cent level.

While the coefficients of correlation presented up to this point indicate the relationships between the various sets of scores and the pupils' ratings, the discussion that follows will indicate the relationship that exists when up to four sets of scores are combined to predict the pupils ratings.

For thirty-three student teachers the scores on the Friendship Rating Scale, cumulative grade averages, student teaching grades, and over-all supervisors' ratings were available. Using a multiple regression technique it was possible to determine the combined correlation between the above four measures and the TPR scores, and, also, the relative contribution of each of the four sets of scores (4). The resulting coefficient of multiple correlation was 0.63. Using an F ratio this was found to be significant at the one per cent level.

By computing both the multiple and partial coefficients of correlation and then obtaining the F ratio it was possible to determine if the elimination of any one set of scores resulted in a statistically significant loss in the prediction of the pupil ratings. Where the over-all supervisors' ratings and the grades obtained in student teaching were eliminated either individually or collectively no statistically significant loss was indicated.

However, when either the Friendship Rating Scale or cumulative grade averages were eliminated a statistically significant loss occurred. The coefficient of multiple correlation using both the Friendship Rating Scale and cumulative grade averages appear to predict the pupils' ratings of student teachers almost as effectively as the combined four scores. The next step was to determine how these two scores might be combined for the most efficient

prediction. A regression equation was computed using only the Friendship Rating Scale and the cumulative grade average and this gave the following equation:

$$Y = 1.76X_1 + 0.67X_2 - 7.32$$

where X_1 represents the scores on the Friendship Rating Scale and X_2 the cumulative grade averages.

The coefficient of correlation between the cumulative grade averages and the TPR scores for the thirty-three cases used in the multiple correlation is considerably higher than the coefficient found for the one hundred and seven cases. Whether this difference is due to chance fluctuation or a real difference between the two samples will need to be determined by further research.

SUMMARY

The problem was to discover the degree that teacher-pupil¹ rapport relates to certain other characteristics of teachers.

The subjects were one hundred and twelve practicing student teachers at the University of Nebraska.

A score to indicate the degree of teacher-pupil rapport (TPR) was found by having each of the pupils choose three teachers he liked best, learned the most from, and obtained help from easiest.

When intelligence scores, supervisors' ratings, and student teaching grades were each correlated with the TPR score, the results were low but positive. Using peer ratings on social characteristics and professors' estimates of teaching ability made three years previously, coefficients as high as 0.5 and 0.6 were found.

A formula for predicting the TPR score was computed using peer ratings and cumulative grade averages.

REFERENCES

- 1) M. P. Barlow, "Adaptation of a Rating Scale Instrument to Sociometrical Evaluation," Unpublished Master's Thesis, University of Nebraska, 1953.
- 2) A. S. Barr, "The Measurement of Teacher Characteristics and Prediction of Teacher Efficiency," *Review of Educational Research*, 22: 1952.
- 3) L. L. Thurstone and Thelma Gwinn, *American Council on Education Psychological Examination for College Freshmen*. Princeton, Cooperative Test Division, 1950.
- 4) J. E. Wert, C. O. Neidt, and J. S. Ahmann, *Statistical Methods in Educational and Psychological Research*. New York, Appleton-Century-Crofts, Inc., 1954.

A LONGITUDINAL COMPARISON OF STUDENT ATTITUDES TOWARD MINORITIES

LEWIS E. ALBRIGHT, ARTHUR D. KIRSCH, C. H. LAWSHE
and H. H. REMMERS

Purdue University

The primary aim of this study was to examine the changes in attitudes toward racial and national groups which have taken place among introductory psychology students at Purdue University over a twenty-year period. Specifically, student-held attitudes toward Germans, Japanese, Jews, and Nazis were sampled five times during this twenty-year span; in 1935 (2), in 1942 (3), in 1945 (4), in 1947 (5) and in 1955. It will be noted that these studies cover the period before, during, and after World War II. In each case the instrument used was Form A of "A Scale for Measuring Attitudes Toward Races and Nationalities" (1) by Grice and edited by Remmers. This is a Thurstone-type attitude scale with scale values ranging from 1 to 11. A high scale value indicates a favorable attitude, whereas a low scale value means an unfavorable attitude toward the race or nationality in question. The indifference point on the scale is 6.

A secondary purpose was to investigate the interrelationships of the attitudes. This was done in order to ascertain whether attitudes toward one minority were significantly related to attitudes toward another.

RESULTS

Table I shows the averages and variability of attitudes found in all five studies. This information is also shown graphically for each group in Figures 1 through 4, the heavy line indicating mean attitude in each case with one standard deviation above and below the mean shaded. As one might expect, there was a severe decline in favorableness of attitude during the war years toward the Japanese and Nazis. The war, however, seemed to have little effect on average attitudes toward Germans and Jews.

It can be seen that there has been a marked increase in mean attitude since 1947 toward each of the groups except the Germans; a very slight decrease in favorableness occurred in that instance.

TABLE I—ATTITUDE DATA BY GROUPS AND YEARS

	1935	1942	1945	1947	1955
Germans					
Mean	8.05	7.98	8.15	8.50	8.48
S.D.	1.17	1.86	1.49	0.98	1.21
S.E. _M	0.11	0.14	0.08	0.05	0.05
Japanese					
Mean	6.11	3.50	3.51	4.69	7.41
S.D.	1.94	1.66	1.19	2.15	1.98
S.E. _M	0.19	0.18	0.13	0.11	0.09
Jews					
Mean	6.30	6.36	6.54	6.40	7.40
S.D.	2.06	2.40	2.43	2.32	1.98
S.E. _M	0.19	0.18	0.13	0.12	0.09
Nazis					
Mean	4.88	3.22	3.02	3.16	4.40
S.D.	2.27	1.62	1.27	1.29	2.51
S.E. _M	0.21	0.12	0.07	0.65	0.11
Minimum N	112	172	335	396	500

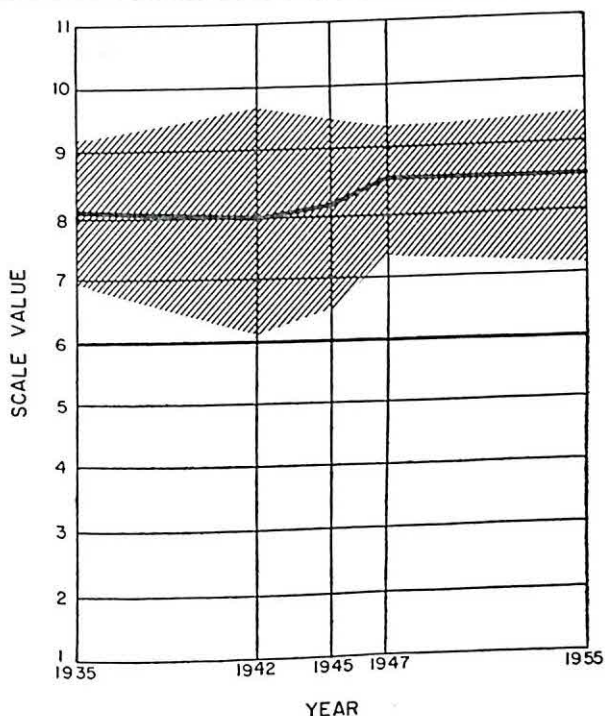


FIGURE 1. Average attitudes toward Germans with one standard deviation above and below the mean shaded.

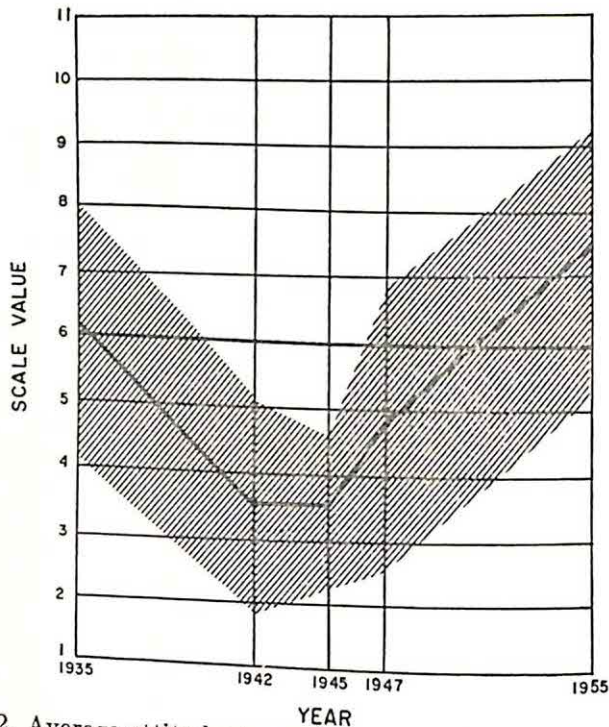


FIGURE 2. Average attitudes toward Japanese with one standard deviation above and below the mean shaded.

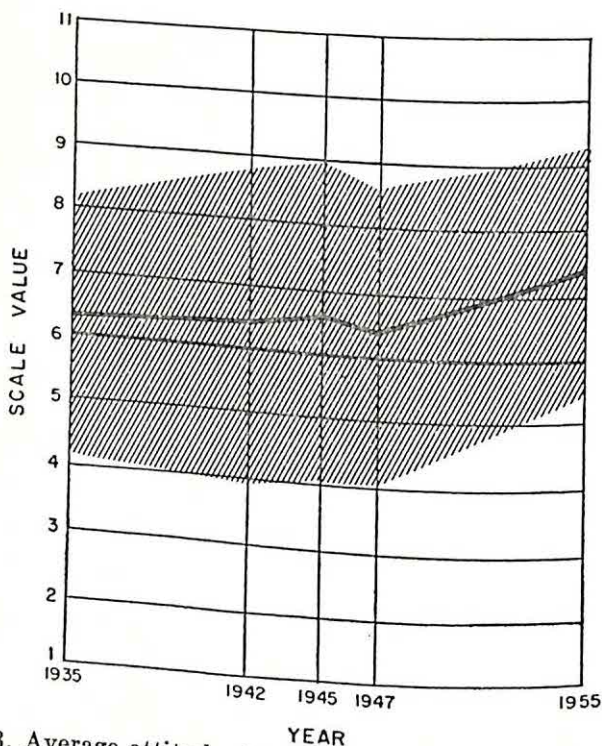


FIGURE 3. Average attitudes toward Jews with one standard deviation above and below the mean shaded.

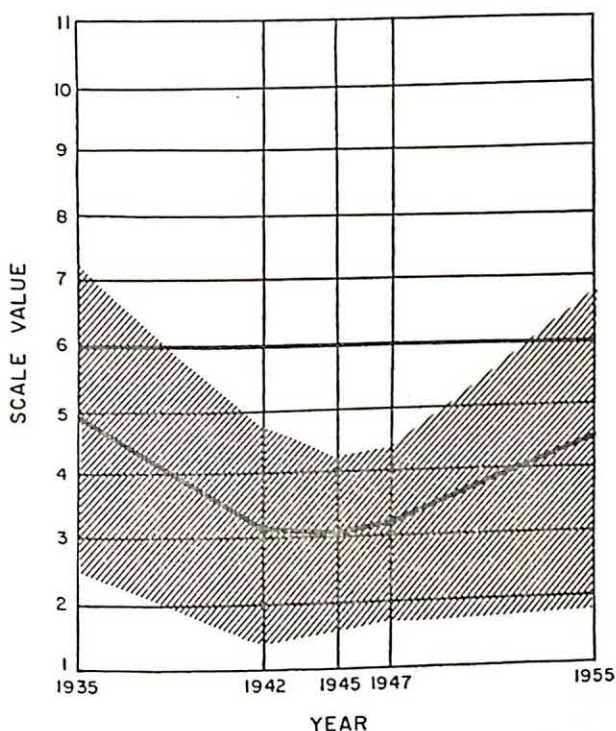


FIGURE 4. Average attitudes toward Nazis with one standard deviation above and below the mean shaded.

However, in all the previous studies including 1947, students felt more favorable toward the Germans than toward any other group. This situation persists in 1955. A test of the significance of each of these changes in average attitude was made, and the results are shown in Table II. The differences between means for the years 1947 and 1955 in the case of Jews, Japanese, and Nazis were significant at the one per cent level. Table II also shows the results of the same significance tests for differences between means from the other preceding studies and those from the 1955 investigation. All of these differences were significant in the direction of increased favorableness in 1955 except the 1935-1955 comparison for Nazis; the result of this comparison is in the direction of decreased favorableness, but not significantly so.

Similar tests of significance were made for the differences between standard deviations to determine the relative variability of attitude from one study to the next. The results of these tests are also shown in Table II. On the basis of these results, one may conclude that

TABLE II—DIFFERENCES BETWEEN MEANS (D_M) AND STANDARD DEVIATIONS (D_{SD}) FROM PRESENT AND PREVIOUS STUDIES, WITH STANDARD ERRORS OF DIFFERENCES AND CRITICAL RATIOS

Groups and Comparisons	Mean			Standard Deviation		
	D_M	SE_D	CR	D_{SD}	SE_D	CR
1947 vs. 1955						
Germans	-0.02	0.07	-0.27	0.23	0.05	4.51**
Japanese	2.72	0.14	19.68**	-0.17	0.10	-1.74
Jews	1.00	0.14	6.96**	-0.34	0.10	-3.29**
Nazis	1.24	0.14	8.95**	1.22	0.11	11.38**
1945 vs. 1955						
Germans	0.33	0.09	3.51**	-0.28	0.07	-4.05**
Japanese	3.90	0.12	32.34**	0.79	0.08	10.13**
Jews	0.86	0.15	5.61**	-0.45	0.11	-3.98**
Nazis	1.38	0.15	9.29**	1.24	0.09	13.33**
1942 vs. 1955						
Germans	0.50	0.12	4.03**	-0.65	0.11	-6.08**
Japanese	3.91	0.17	23.30**	0.32	0.11	2.93**
Jews	1.04	0.18	5.62**	-0.42	0.14	-2.93**
Nazis	1.18	0.20	5.77**	0.89	0.12	7.57**
1935 vs. 1955						
Germans	0.43	0.12	3.44**	0.04	0.09	0.46
Japanese	1.30	0.25	5.28**	0.04	0.14	0.28
Jews	1.19	0.25	4.41**	-0.08	0.15	-0.53
Nazis	-0.48	0.31	-1.55	0.24	0.17	1.42

** Significant at the one per cent level.

attitudes have become significantly more heterogeneous toward Germans and Nazis since 1947, but significantly less variable toward Jews in the same period. Compared to 1955, attitudes toward Germans and Jews in 1942 and 1945 were significantly more variable, but were significantly less variable toward Nazis and Japanese in the same years. There were no statistically significant changes in variability of attitudes between the 1935 study and the present.

When the interrelationships of the attitudes were plotted for males and females,¹ it was found that most of the regressions tended toward curvilinearity. In fact, the correlation ratios were found to be significant ($p < 0.01$) for all but three relationships; Nazis vs.

¹ A table showing the intercorrelations of the attitudes and a breakdown of the respondents by sex has been deposited with the American Documentation Institute, 1719 N St. N.W., Washington 6, D. C.

Germans for males and females and Nazis vs. Jews for males only. When a test for the significance of linear correlation was applied, five significant F-ratios were obtained; i.e., Nazis vs. Japanese for females, Japanese vs. Jews and Germans vs. Japanese for both males and females. All relationships were significant at the one per cent level except that of Germans vs. Japanese for females, which was significant at the five per cent level. These same five relationships were also found to depart significantly from linearity when tested, indicating the increased prediction effectiveness to be obtained from a curved regression line. All of these F-ratios were significant at the one per cent level except that for Japanese vs. Jews (males only), significant at the five per cent level.

DISCUSSION

Probably the most outstanding finding with respect to the present study is the more favorable attitude shown by the 1955 respondents toward the Japanese, Nazis and Jews. This result might be expected in the case of the Japanese and Nazis as memories of World War II become less distinct. Perhaps the increased variability of attitudes toward both of these groups since the war indicates that some people tend to "forget and forgive," whereas others do not. Further support for this view was the finding that a few students in the present study seemed to have trouble remembering who Nazis were. One student wrote on his paper, "I have never met a Nazi, so I cannot answer."

An alternate hypothesis, the "general tolerance factor" postulated by Remmers (4), might also be advanced. This "general tolerance factor" would operate "... in such a way that a given individual is likely to be more favorable toward any racial or national group as compared with some other individual, even though the group averages for the different national and racial groups are widely separated" (4). Remmers postulated this factor on the basis of a low, but significant, correlation between attitudes toward Jews and Nazis. Although this Jew vs. Nazi relationship was found to be significant only for the present female population, the general tolerance factor is still inferable from the other positive intercorrelations found in this study.

Still a third explanation might be that the word "Nazi" has lost its impact for the 1955 sample and has become vague in meaning, a semantic blurb with only hazy connotations.

The increased favorableness toward Jews in 1955 is somewhat surprising considering the remarkable stability of mean attitudes toward Jews exhibited in the previous studies. Two possible explanations immediately suggest themselves. One possibility is that students are becoming more tolerant toward Jews as a result of some psychological or social phenomenon, e.g., brotherhood campaigns, increased contact with Jews, pro-Semitic literature, etc. The other explanation is that there may be a higher percentage of Jewish students in introductory psychology in 1955 than there was in the years of the previous studies. Evidence pertinent to this latter hypothesis was not obtainable. It is recognized that some combination of these two possibilities could also be true.

The significantly smaller standard deviation found for Jews in this investigation indicates that students feel more alike in their favorable attitude toward Jews than did three preceding groups of respondents in their less favorable attitudes. However, since the standard deviation remains relatively large, one can conclude that there is still a wide variety of opinion present.

The fact that Germans have consistently ranked highest in favorableness of any of the four groups over the twenty-year period bears some comment. One probable reason for this finding may be that many of the students who come to a Midwestern institution like Purdue University are descendants of the rather large German and North European population which emigrated to the Midwest.

It is certain that, whatever the reason, Germans are not now as controversial a group in the opinion of the respondents as are Jews, Japanese, or Nazis. During the war years, however, one notes the larger standard deviation for Germans, even though the mean attitude remained highly favorable. Evidently most people made a clear distinction between Germans and Nazis, but there was more heterogeneity of attitude than either before or since the war. Since 1947 there has been virtually no change in mean attitude toward this group, although variability has increased significantly. A tentative hypothesis which could be advanced concerning the cause of this latter finding is the division of power in Germany. Students in 1955 might have one attitude toward East Germans and another toward West Germans. Since they were asked their opinion of "Germans" collectively, more variability of responses might be expected than was the case some years ago.

SUMMARY AND CONCLUSIONS

Five measures of attitudes toward Germans, Japanese, Jews and Nazis were made over a twenty-year period. Presumably similar samples of introductory psychology students at Purdue University served as subjects in the years 1935, 1942, 1945, 1947, and 1955. The results support these conclusions with regard to the present study:

(1) A significant increase in favorableness of attitude has occurred toward Japanese, Jews, and Nazis since 1947. This increase was accompanied by significantly greater homogeneity of attitudes toward Jews, and significantly greater heterogeneity of attitude toward Nazis.

(2) The Japanese and Jews were rated significantly more favorable than in any of the previous studies.

(3) No significant change was found in mean attitude toward Germans since 1947, although greater heterogeneity of attitude did occur.

(4) In this study, as in each of the previous investigations, Germans were ranked highest of the four groups in favorableness.

(5) Partial evidence still supports Remmers' "general tolerance factor."

(6) The implications of the study concerning intergroup tensions and the possibilities of discrimination on the basis of race, creed or national origin in employment, housing and social relations generally are obvious and need no explication here.

REFERENCES

- 1) H. H. Remmers, et al., "Studies in Attitudes—a Contribution to Social Psychological Research Methods," *Studies in Higher Education*, 26, *Bulletin Purdue University*, 35, 1934.
- 2) H. H. Remmers, and C. L. Morgan, "Changing Attitudes toward a Racial Group," *Further Studies in Attitudes, Series II, Studies in Higher Education*, 31, *Bulletin Purdue University*, 37: 109-114, 1936.
- 3) H. H. Remmers, "Attitudes toward Germans, Japanese, Jews, and Nazis as Affected by the War," *School & Society*, 57: 138-140, 1943.
- 4) H. H. Remmers, "Changes in Attitudes toward Germans, Japanese, Jews, and Nazis as Affected by the War," *School & Society*, 63: 118-119, 1946.
- 5) H. H. Remmers and W. F. Wood, "Changes in Attitudes toward Germans, Japanese, Jews, and Nazis," *School & Society*, 65: 484-487, 1947.

BOOK REVIEWS

TRUMAN M. PIERCE, ET AL. *White and Negro Schools in the South*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1955, pp. 338.

The co-authors with Dean Pierce in this extensive and very timely study are James B. Kincheloe, R. Edgar Moore, Galen N. Drewry and Bennie E. Carmichael, all of whom are especially experienced and informed in the field of education in the South. For the purpose of this investigation, the South is defined as the states of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas and Virginia.

In the summer of 1953, The Fund for the Advancement of Education undertook to finance the research. "A large staff of Southerners who were familiar with the educational scene in the region was quickly assembled, and the most searching study of education on a regional level ever undertaken was launched The staff of the Southern States Coöperative Program in Educational Administration accepted the responsibility for conducting the research on the dual school systems Each agency declined the rôle of a protagonist in the segregation issue and sought only to provide, in as objective a fashion as possible, data essential to those in whose hands policy determination is lodged" (p. vi).

The report is divided into three sections. Section 1 deals with the issues—their setting and development—in three chapters: Public Education and American Society, The Historical Context of Biracial Education in the South, and The Organization and Administration of Biracial Education.

Section 2 discusses the public schools—problems and progress—in six chapters: School-Age Population, Public School Enrollment and Attendance, Public School Revenues and Related Factors, Patterns of Public School Expenditures, Public School Personnel, Public School Facilities, and The Educational Problem.

Section 3 deals with the issues redefined—looking ahead—in two chapters: Taking Stock and Facing the Issues.

Appendix A discusses the sources of data. Appendix B gives tables and charts of additional statistical data. A four-page index completes the book.

Although the historical introductory chapters in Section 1 and the discussions will be of special interest to many readers, especially those who are not acquainted with the development of the situation in the South, the greatest value attaches to the extensive statements of factual data found in Section 2. The facts for each of the thirteen Southern States are given separately and with comparisons for white and Negro schools. These comparisons include revenues and expenditures; school personnel, teacher load, teacher training and teacher's salaries; value and size of buildings, equipment, use of facilities, school library service, opportunities for vocational education, and school lunch service.

These data are presented in ninety-nine tables and twenty-five charts. This Section represents long, extensive and careful analytical work relating to facts for the region as a whole, for the states separately, and for whites and Negroes in comparison with each other.

Discussion in the volume aims at avoiding prejudice and trying to follow the facts presented and also presenting general principles upon which decisions might appropriately be made. The authors do not attempt to give final solutions to the problems of segregation. They especially emphasize the need for scientific facts upon which to base wise judgments, and the scientific attitude for dealing with all of the problems.

This scientific procedure seems to have been followed in making the report with the exception of the psychological area: Here, except for some reference to attitudes and beliefs, it seemed to be sufficient to quote Gunnar Myrdal, sociologist. The reader is left with the impression that a supplementary report upon psychological information, presented by psychologists, might be very appropriate. Many data are available.

Emphasis is laid not only on the need for the use of the scientific spirit and method, but for understanding as thoroughly as possible all the facts involved, and for wise and law-abiding attitudes in making decisions and in final action.

The volume is a valuable contribution at the present time for those who desire to find a comprehensive statement of concrete details in relation to biracial education in the South, whether or not it may appear to some that departure from strict scientific procedure and objectivity may have crept in on occasion. It will probably be found to be one of the least prejudiced accounts of

the subject upon which so many uninformed and emotional statements can be found.

A. S. EDWARDS

The University of Georgia

ALBERT K. COHEN. *Delinquent Boys: The Culture of the Gang*. Glencoe, Illinois: The Free Press, 1955, pp. 202.

The growing understanding of the importance of culture and of cultures finds expression in this book which leads to what has been called "a distinct contribution to the study of delinquency and crime." The point of view is not that of the delinquent, but of the "delinquent subculture." Most significant are the different cultures, models, techniques, etc., of different groups.

In five chapters the author discusses An Unsolved Problem in Juvenile Delinquency, Facts the Theory Must Fit, A General Theory of Subcultures, Growing Up in a Class System, and A Delinquent Solution. The book ends with a Conclusion and with notes and an index.

The discussions include the nature and content of the delinquent subculture, pressures toward conformity, some subcultural attempts at solutions of problems, cultural settings, personality differences, delinquency control, with special emphasis upon the problems and situations found among middle- and working-class children.

With some variation the problem or problems of the book are stated (p. 109) as "What is it about the structure of American society that produces, in certain sectors of that society, a subculture of a certain distinctive content?" On p. 148, we find that "The problem that this book tries to answer is this: Why does a particular subculture, with a certain distinctive content, have its principal locus in a certain role sector in our society?" More briefly (on p. 19) the "task of this book is to account for the delinquent subculture itself."

Careful reading will discover the heart of the problems discussed and be properly rewarded for a thoughtful consideration of the matters that are presented. Emphasis is placed upon the social world outside the family but not to the loss of emphasis upon the conditions within the family. Special stress is laid upon the point of view of group solving of problems. Not only is it a question,

Why is this boy delinquent? but also, Why do we have this delinquent subculture?

Into what the author considers to be "a curious gap in delinquency theory," he throws his efforts to bring out better insight concerning a subject that has had too little consideration and understanding.

The problems dealt with are primarily those of boys, but brief consideration of delinquency relating to girls is included. The classes included are the working- and middle-class boys. Some existing theories are considered and thought to be inadequate; the author presents what he believes to be more nearly true and useful and makes suggestions for future research.

Cohen modestly evaluates his book as a "prolegomenon and a signpost to further research," and hopes that it gives "tentative solution to a major theoretical problem." After summarizing certain questions, he says: "Perhaps it is sufficient justification of the present work to have put these questions bluntly."

For the reader not acquainted with the relevant literature, the Notes to Chapters (pp. 181-195) will be helpful. A useful index (pp. 197-202) concludes the book. It is a constructive contribution to this field.

A. S. EDWARDS

The University of Georgia

RUTH KOTINSKY AND HELEN L. WITMER, Editors. *Community Programs for Mental Health*. Cambridge: Harvard University Press, 1955, pp. xix, 358.

This is one of the books published for The Commonwealth Fund. Two specially prepared editors were selected, and also eight competent contributors, namely, Barbara Biber, H. E. Chamberlain, Sol W. Ginsburg, Robert R. Holt, Louisa P. Howe, Marie Jahoda, Elizabeth deSchweinitz, and Edith Miller Tufts. In their foreword the editors write that these papers "are to be regarded as depicting the current state of mental health promotion work in the United States, with some suggestions for directions to be taken if the aim of the work is to be clarified and fulfilled."

Part 1 deals with Theory. Here Ginsburg gives appropriate warning against the dangers of inadequate theory and attempts to

develop an appropriate attitude towards theory. Mental health needs more satisfactory definition and more adequate theory.

Part 2 deals with Practice. Tufts discusses difficulties of delimiting the field, promotion, prevention and education, and indicates the great variety of programs, their organizations, sponsorships, groups served, financial problems, etc.

The main part of the discussion by Chamberlain and deSchweinitz deals with the details of nine programs for mental health that have been developed in the United States. They fall in four main categories: (a) Programs ranging in focus from the concentrated to the diffuse; (b) Programs that keep services and research workably geared; (c) those that are not psychiatrically oriented; and (d) a special program of statewide consultation. Excellent discussions of the details of these programs are given, including origins, sponsorship, staff, policies, objectives, programs, and some indication of financial support and evaluation of results.

Biber discusses schooling as an influence in developing healthy personality and gives an analysis of three school projects.

In Part 3, under evaluation, Howe discusses the problems of evaluation, objectives and underlying theories, assumptions and the gauging of community mental health. Jahoda deals with the social psychology of mental health, her study being the reprint of a paper originally appearing in a symposium on the Healthy Personality.

The appendix, by Holt, is an outline of the problems in the making of sample surveys. Here he details many usable suggestions for planning a survey, choosing a design for the study, selecting the sample, developing the questionnaire, administering the survey, analyzing results and writing the report, etc.

This volume is especially valuable in giving information about what is being done in this country for mental health, and indicating, for those who wish to do their best for this urgently needed work, what can be done. It will be found to be a very useful addition to the growing literature of good books on the subject which is as yet too little known and too inadequately understood.

The University of Georgia

A. S. EDWARDS

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

November, 1956

Number 7

SEX DIFFERENCES IN THE DISTRIBUTION OF TEACHER APPROVAL AND DISAPPROVAL AMONG SIXTH-GRADE CHILDREN

WILLIAM J. MEYER and GEORGE G. THOMPSON

Psychology Department, Syracuse University

This study was designed to investigate the relative frequency of women teachers' approval and disapproval evaluations of sixth-grade male as contrasted with female pupils. The relevant data for this study were obtained by means of two independent techniques: thirty hours of direct observation of teacher-pupil interactions in each of three classrooms; and, the use of a modification of the "Guess Who?" technique to determine if the children themselves were aware of any sex differences in their teachers' approval and disapproval evaluations.

There is considerable agreement among psychologists that the use of approval by the teacher results in better learning and probably in better over-all adjustment (14, 18). Some studies (19, 24) have shown that personal maladjustments in teachers have deleterious effects on the adjustment level of the children in their classes.

In a series of studies by Anderson *et al.* (1-3), using a sample of kindergarten-age children, the data indicate that teachers typically use statements of a dominative nature in their interactions with the children in their classes. Anderson further reports that the teachers in his study tended to levy most of their dominative and/or integrative overtures on only a few pupils to the relative neglect of the other children in the classroom. Further evidence of this nature is reported in a study by deGroat and Thompson using four sixth-grade classrooms (10). In addition to reporting inequities in teacher approval and disapproval they also found that teachers give more praise to the youngsters who are brighter, better adjusted and higher achievers. The more poorly adjusted

and the duller children were observed by these investigators to receive more disapproval from their teachers.

The purpose of the present investigation is to shed more light on the ways in which teachers respond toward the pupils in their classrooms. Extensive research findings have been reported in the literature (11, 12, 20, 21, 28) which consistently show that boys are more aggressive and generally more "unmanageable" than girls. It is our hypothesis that this "masculine" behavior will result in male pupils receiving a larger number of dominative, or punitive, contacts than girls from their teacher, who is usually a woman from the middle socio-economic stratum of our society. That is, we feel that the behavior of boys in the typical classroom is of such a nature as to make it less acceptable to teachers who probably attempt to perpetuate certain middle-class standards of what "good" classroom behavior should be. We believe that girls usually display behavior more in conformity to the standards perceived as "good" by the average elementary school teacher and will therefore receive fewer disapproval contacts and more approval contacts from their teachers.

Assuming that the above hypotheses are supported by the data we would also predict that children of elementary school age will recognize, and take for granted, that boys receive more disapproval and blame from their teachers than girls.

EXPERIMENTAL PROCEDURE

In order to test the hypothesis that boys receive a larger number of dominative, or disapproval, evaluations from their teachers than do girls, teacher-pupil interaction within three sixth-grade classrooms were recorded for a total sample of thirty hours per classroom. These time samples of classroom behavior were spread over an entire school year. Among other things being studied, interactions between teachers and pupils were classified into two categories: (a) praise contacts (teacher initiated interactions with a child in which she verbally expressed approval of some behavior which the child had displayed), and (b) blame contacts (teacher initiated interactions with a child in which she verbally expressed disapproval for some bit of behavior which the child had displayed). Observer agreement for the praise classification ranged from eighty-four to one hundred per cent with a median of approximately ninety-two per cent. Observer agreement for the blame classifica-

tion ranged from fifty-seven to one hundred per cent, with a median of approximately ninety-three per cent.

In an attempt to cast some light on children's perceptions of any sex differences in teacher disapproval, a modified "Guess Who?" approach was employed. The "Guess Who?" approach used in this study required each child to nominate fellow class members for a number of situations in which children are receiving approval or disapproval from their teacher for some behavior. (See deGroat and Thompson (10) for a more complete description of these scales and information about their reliabilities.) The behavior descriptions were selected on the basis of their familiarity to children and contain a fairly representative sample of situations in which children typically receive either approval or disapproval from their teachers. Each child was required to list the names of four of his classmates whom he thought fitted each of the behavior descriptions most adequately.

RESULTS

Fisher's *t* test was used to determine the reliability of the obtained sex differences.¹ The difference between disapproval contacts received by boys and by girls from their teachers was statistically significant in each of the three classrooms. As predicted, the boys received the larger number of disapproval contacts. These differences may be interpreted according to our hypothesis as supporting the notion that teachers are responding with counter-aggression to the greater expression of aggression by boys. The results obtained in analyzing the teachers' praise contacts with boys and girls are presented in Table II. The only statistically significant differences obtained for this variable was in school B. However the boys received more praise than the girls in each of the classrooms. It may be that the teachers are attempting to reinforce any positive behavior that the boys may display. Or this tendency to praise boys more than girls may reflect compensatory behavior for guilt feelings created in the teacher by her excessive aggressive-

¹This test assumes that the samples being compared are homogeneous with respect to their variances. Frequently this assumption had to be rejected in some of the group comparisons. In such cases a more conservative test of significance was used which makes some allowance in the error term for heterogeneity of variance. This technique is presented in detail in Cochran and Cox (6).

TABLE I—SEX DIFFERENCES IN FREQUENCY OF TEACHERS' DISAPPROVAL CONTACTS

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	11.10	2.67	10.75	2.79	10.06	1.44
S.D.	7.62	2.18	9.27	2.42	14.42	1.59
t	3.20**		3.11**		2.37*	
F	12.22**		14.67**		130.36**	
t. _{.01} †	3.30		3.10		2.92	

* Significant at the five per cent level.

** Significant at the one per cent level.

† See footnote 2.

TABLE II—SEX DIFFERENCES IN FREQUENCY OF TEACHERS' APPROVAL CONTACTS

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	9.90	9.67	10.50	5.50	3.71	2.69
S.D.	6.10	7.65	5.33	2.53	2.95	1.99
t	0.074		2.50*		1.15	
F	1.57		4.43**		2.19	
t. _{.05} †	—		2.19		—	

* Significant at the five per cent level.

** Significant at the one per cent level.

† See footnote 2.

ness towards boys. Either interpretation, or any one of the several others that could be offered, is highly speculative.

The data presented above are based on the extensive observations of an objective observer who played no functional rôle in the classrooms. The data presented in the following section reflect the teachers' approval and disapproval contacts as viewed by their pupils.

"GUESS WHO?" DATA

Analysis of the "Guess Who?" data was performed along the same lines as the data obtained by direct observation. A comparison

TABLE III—SEX DIFFERENCES IN CHILDREN'S NOMINATIONS FOR TEACHER DISAPPROVAL

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	21.60	5.33	42.33	9.71	33.82	5.18
S.D.	13.33	5.68	43.06	8.13	55.87	8.26
t	3.39**		2.79**		2.03	
F	5.51**		28.07**		19.21**	
t. _{.01} †	3.26		3.10		—	

** Significant at the one per cent level.

† See footnote 2.

of the pupils' nominations of their peers on the disapproval items revealed statistically significant differences between boys and girls for two of the three schools. This can be interpreted as showing that the boys are viewed by the girls as well as by their male peers as being involved in more situations which evoke disapproval from their teachers.

Analysis of the children's responses to the items related to teacher approval produced no significant differences between boys and girls.

A final analysis of the "Guess Who?" data was performed in an attempt to determine how boys as contrasted with girls perceived the teacher's approval and disapproval biases. The choices

TABLE IV—SEX DIFFERENCES IN CHILDREN'S NOMINATIONS FOR TEACHER APPROVAL

	Classroom A		Classroom B		Classroom C	
	Boys	Girls	Boys	Girls	Boys	Girls
N	10	9	12	14	17	16
Mean	11.00	34.33	23.58	31.42	23.71	21.60
S.D.	12.93	32.64	18.51	25.08	35.75	25.70
t	2.09		0.856		0.242	
F	6.376**		1.835		19.35**	
t. _{.01} †	—		—		—	

** Significant at the one per cent level.

† See footnote 2.

TABLE V—CHOICES MADE BY BOYS AND BY GIRLS
ON TEACHER DISAPPROVAL ITEMS

	Classroom A				Classroom B				Classroom C			
	Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing	
	B	G	B	G	B	G	B	G	B	G	B	G
N	8	9	8	9	14	12	14	12	17	16	17	16
%	89.77	23.23	83.09	16.91	88.37	11.63	73.05	26.95	88.70	11.20	82.21	17.79
CR	5.60**		3.64**		6.10**		2.65**		7.03**		4.84**	

** Significant at the one per cent level.

made by the boys and by the girls for the approval and disapproval items were separately analyzed. It seemed unreasonable to use the *t* test in this situation because of the unequal numbers of boys and girls in the classroom. Therefore the groups were equated by converting the frequencies of nominations to percentages and working with percentage differences.

The results of the statistical analysis of boys' nominations on the disapproval items show that boys respond as if they usually received more blame from teachers than do girls. It would appear that boys are quite sensitive to the disapproval of their teachers. Table V shows that the girls also respond as if boys receive more teacher disapproval.

There is little consistency in the nominations made by the boys for the praise items. In schools B and C the boys react as though

TABLE VI—CHOICES MADE BY BOYS AND BY GIRLS
ON TEACHER APPROVAL ITEMS

	Classroom A				Classroom B				Classroom C			
	Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing		Boys Choosing		Girls Choosing	
	B	G	B	G	B	G	B	G	B	G	B	G
N	8	9	8	9	14	12	14	12	17	16	17	16
%	36.98	63.01	34.07	65.92	56.95	43.04	24.58	75.42	57.56	42.43	46.22	53.77
CR	1.11		1.38		0.72		3.01**		0.88		0.44	

** Significant at the one per cent level.

they typically receive more praise than girls, although this difference is not statistically significant. In contrast to the boys' responses, the girls feel that they receive more praise, particularly in school B where the difference is statistically significant. These results might be interpreted as meaning that children fail to recognize any definite dichotomy in the teacher's distribution of praise contacts.

DISCUSSION

The general findings of this study support the hypothesis that the male pupil receives reliably more blame from his teacher than the female pupil. Moreover, the boys recognize that they are the recipients of a higher incidence of teacher disapproval. We feel that these data lend indirect support to the notion that "masculine" behavior is not tolerated by the typical teacher who in turn attempts to inhibit such behavior by means of punishment.

Davis and Havighurst (8) have discussed at length the divergence of cultural mores between lower-class children and their middle-class teachers. Their work may best be summarized in the assertion that the goals defined by the middle-class teacher do not receive reinforcement from the lower-class child's peer group or from his family. Teacher initiation of punishment for "misbehavior" only serves to reinforce an already existing dislike for school and further leads to peer group reinforcement. A similar (but by no means identical) interpretation appears relevant to the present discussion. Our society's definitions of acceptable male and female behavior are divergent particularly with respect to aggression. For example Radke (20) in her monograph on the relationship of parental authority to child behavior reports that the fathers in her sample felt that aggressive, assertive behavior on the part of boys was less undesirable than the identical behavior in girls (and in many cases was deemed highly desirable). The mothers felt that aggression was unacceptable behavior in either sex but in general they were in agreement that aggressive behavior is more unacceptable in girls. In another study specifically related to the notion that aggressive behavior is more unacceptable in the female culture is a study by Sears *et al.* (22). These writers predicted that in father-absent homes, wherein the child is brought up by the mother, boys would be less aggressive than in father-present homes in which the boy models his behavior after the father. The results

of their study support the "sex-typing" hypothesis as presented above. Bach (4) reports similar evidence in support of the "sex-typing" hypothesis.² Apparently the social mores of the typical female teacher, at least with respect to aggressive, assertive behavior, are in sharp contrast to the behavioral tendencies of the typical male youngsters. The behavioral tendencies of the female child are, however, in close agreement with those of her teachers. We feel that the above generalization accounts to a high degree for the data reported in this study. Our argument becomes somewhat stronger when the work of Wickman (29) and a follow-up study by Mitchell (17) are included in the discussion. These investigators found that teachers perceive aggressive nonconforming behavior as more serious than withdrawal behavior. More recently Kaplan (15) has reported that the aggressive child was deemed annoying to almost three-quarters of the teachers in his sample. The present investigation suggests that perhaps teachers react to the aggressive behavior of children with counter-aggression, a vicious circle for both pupil and teacher.

Consistent with the above interpretation is the larger amount of variation found among the male pupils as contrasted with the female pupils. In a culture such as ours in which the father is away from the home during most of the child's waking hours (and in some instances pays only cursory attention to the youngster when at home), it appears obvious that both the male and female child are more directly influenced by the mother. Many boys, however, will be influenced more by their fathers and peer culture than by the mother because of identification with the masculine rôle in our culture. Our belief is that these more "masculine" boys are the ones who receive the greater share of teacher disapproval. Such an interpretation appears consistent with the work of Sears (22) and Bach (4).

The foregoing discussion has certain implications for the student of child development and education. If our interpretation of the teacher and male-pupil relationship is accurate, then the fact that boys dislike school more than girls is understandable. The daily punishment received by the boy for behavior he really does not consider "bad" must certainly be anxiety producing. If the

² Though there is insufficient evidence at this time the present writers feel that the factor of innate sex differences in aggressive tendencies should not be overlooked. See Beach (5) for suggestive findings.

anxiety created in the school situation becomes sufficiently intense, it seems reasonable that tension reduction can be achieved by means of avoiding school. It is known that more boys leave school at an earlier age than girls (26).

Perhaps of even more importance is the effect of this teacher-disapproval generated anxiety on the general personality adjustment of male pupils. It is unfortunate that we do not have evidence on the changes in adjustment level of the children in our sample, but studies by Ojemann and Wilkinson (18) and others indicate that consistent teacher dominance has deleterious effects on the adjustment of children. We can only speculate as to the nature of these adjustment problems but such behavioral manifestations as nervousness, withdrawal and lack of self-confidence are a few of the known symptoms.

We feel that the consistent trends in our findings imply that teachers' negative attitudes towards their male pupils arise from a lack of appreciation for the term "normal" male child. In our culture, aggressive outgoing behavior is as normal in the male as quiescent nonassertive behavior is in the female. The teacher who attempts to thwart this behavior by means of threats and punishment can only meet with frustration since the boy is confronted with a conflicting social code. A more reasonable plan to follow would seem to be one in which the excess energy and tensions of the male child could be discharged on some constructive activity. Planned physical education classes will do much to dissipate aggressive needs in a socially acceptable manner. Perhaps most important of all, however, is the knowledge that some degree of aggressive behavior is a normal part of development in both boys and girls and should be treated not as a personal threat to the teacher but as sign of "normal" social and personality development.

SUMMARY

The purpose of this study was to investigate sex differences in teacher distribution of approval and disapproval among three sixth-grade classrooms. Data relevant to the children's perceptions of their teachers' attitudes towards boys and girls were also collected. Using the discrepancies in attitude between males and females in our culture toward aggressive behavior as the basic underlying variable, the hypothesis was offered that boys, who

Bureau Ednl. Res. Research

DAVID H.

COLLEGE

Dated

are more aggressive and nonconforming than girls, would receive more disapproval contacts from their teachers than girls. Girls being quiescent and more conforming than boys would as a consequence receive more approval from their teachers than boys. We further hypothesized that both boys and girls will be aware of the differences in their teachers' attitudes towards them.

In order to test the foregoing hypothesis three sixth-grade teachers and their pupils were directly observed for a total of thirty hours per classroom. All teacher initiated contacts of an approval or disapproval nature were recorded. The measurement of the children's perceptions of teacher attitude was accomplished by means of a variation of the "Guess Who?" technique. The pupils were asked to list the names of four students who best fitted a series of statements of a teacher approval nature and of disapproval nature. Analysis was made of the number of children of each sex chosen for the approval items and for the disapproval items.

Statistical analysis of the data clearly supports our hypothesis with respect to male pupils. In all three schools the boys received reliably more disapproval from their teachers than the girls. We also found that both the boys and the girls nominated more boys for the disapproval items than girls. This difference was statistically reliable. With respect to the second hypothesis concerning girls, the data did not yield any clear-cut differences. If any trend was present it was in a direction opposite to that predicted. These results indicated that the teachers in our sample tended to have fewer contacts with the girls in their classrooms.

The results of this investigation were interpreted as being consistent with the notion of a sex difference in attitude towards aggressive behavior. The conclusion was drawn that teachers attempt to "socialize" the male child by means of dominative counter-aggressive behavior. The negative consequences of this situation for the child are discussed.

REFERENCES

- 1) H. H. Anderson and H. M. Brewer, "Dominative and Socially Integrative Behavior of Kindergarten Teachers," *Applied Psychology Monograph No. 6*. Stanford, Calif., Stanford University Press, 1945.
- 2) H. H. Anderson and J. E. Brewer, "Effects of Teachers' Dominative and Integrative Contacts on Children's Classroom Behavior," *Applied*

Psychology Monograph No. 8. Stanford, Calif., Stanford University Press, 1946.

3) H. H. Anderson, J. E. Brewer, and M. F. Reed, "Studies of Teachers' Classroom Personalities: III. Follow-up studies of the effects of dominative and integrative contacts on children's behavior," *Applied Psychology Monograph No. 11*. Stanford, Calif., Stanford University Press, 1946.

4) G. R. Bach, "Father-Fantasies and Father-Typing in Father Separated Children," *Child Development*, 17: 63-80, 1946.

5) F. A. Beach, *Hormones and Behavior*. New York, Paul B. Hoeber, 1948.

6) W. G. Cochran and G. M. Cox, *Experimental Designs*. New York, Wiley & Sons, Inc., 1950.

7) W. A. Davis, *Social-Class Influences upon Learning*. Cambridge, Harvard University Press, 1948.

8) W. A. Davis and R. J. Havighurst, *Father of the Man: How Your Child Gets His Personality*. Boston, Houghton Mifflin Co., 1947.

9) W. A. Davis and J. Dollard, *Children of Bondage*. Washington, D. C., American Council on Education, 1940.

10) A. F. deGroat and G. G. Thompson, "A Study of the Distribution of Teacher Approval and Disapproval among Sixth-Grade Pupils," *Journal of Experimental Education*, 18: 57-75, 1949.

11) F. L. Goodenough, "Anger in Young Children," *Child Welfare Monograph No. 9*, pp. xiii + 278. Minneapolis, University of Minnesota Press, 1931.

12) M. L. Hayes, "A Study of the Classroom Disturbances of Eighth Grade Boys and Girls," *Teachers College Contributions to Education*, 871: 139, 1943.

13) E. Hollenberg and M. Sperry, "Some Antecedents of Aggression and Effects of Frustration on Doll Play," *Personality*, 1: 32-43, 1951.

14) E. Hurlock, "The Value of Praise and Reproof as Incentives for Children," *Archives of Psychology*, 11: No. 71, 1924.

15) L. Kaplan, "The Annoyances of Elementary School Teachers," *Journal of Educational Research*, 45: 649-665, 1951-52.

16) C. Leeds and W. Cook, "The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes," *Journal Experimental Education*, 16: 149-159, 1947.

17) J. C. Mitchell, "A Study of Teachers' and of Mental-Hygienists' Ratings of Certain Behavior Problems of Children," *Journal of Educational Research*, 36: 292-307, 1942.

18) R. Ojemann and F. R. Wilkinson, "The Effect on Pupil Growth of an Increase in Teacher's Understanding of Pupil Behavior," *Journal of Experimental Education*, 8: 143-147, 1939.

19) W. Olson and M. Wilkinson, "Teacher Personality as Revealed by the Amount and Kind of Verbal Direction Used in Behavior Control," *Educational Administration and Supervision*, 24: 81-93, 1938.

20) M. J. Radke, "The Relation of Parental Authority to Children's Behavior and Attitudes," *Child Welfare Monograph No. 22*, Minneapolis, University of Minnesota Press, 1946.

21) P. S. Sears, "Doll Play Aggression in Normal Young Children: Influence of Sex, Age, Sibling Status, Father's Absence," *Psychological Monographs*, 65: 42, 1951.

22) R. R. Sears, M. H. Pintler, and P. Sears, "Effect of Father Separation on Preschool Children's Doll Play Aggression," *Child Development*, 17: 219-243, 1946.

23) R. Sears, J. Whiting, V. Nowlis, and P. Sears, "Some Child-rearing Antecedents of Aggression and Dependency in Young Children," *Genetic Psychology Monographs*, 47: 135-234, 1953.

24) W. Snyder, "Do Teachers Cause Maladjustment?," *Journal Exceptional Children*, 14: 40-46, 1947.

25) R. T. Sollenberger, "Some Relationships between the Urinary Excretion of Male Hormone by Maturing Boys and Their Expressed Interests and Attitudes," *Journal of Psychology*, 9: 179-189, 1940.

26) S. Tenenbaum, "Uncontrolled Expressions of Children's Attitudes toward School," *Elementary School Journal*, 40: 670-768, 1939-40.

27) S. Tiedeman, "A Study of Pupil-Teacher Relationship," *Journal of Educational Research*, 35: 657-664, 1942.

28) R. D. Tuddenham, "Studies in Reputation: I. Sex and Grade Differences," *Psychology Monographs*, 66: 58, 1952.

29) E. K. Wickman, *Teachers and Behavior Problems*. New York, Commonwealth Fund, 1938.

30) J. Withall, "The Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms," *Journal of Experimental Education*, 17: 347-361, 1949.

A COMPARISON OF THE FACTORIAL STRUCTURE OF COGNITIVE FUNCTIONS FOR A HIGH AND LOW STATUS GROUP¹

JAMES V. MITCHELL, JR.

Miami University

Comparative studies dealing with the similarities and differences between the structure of mental abilities for different groups may be of value in several ways. For those who are interested in the efficacy and consistency of a particular method of factor analysis, such comparisons become tests of the "stability" or invariance of a factorial solution for groups of dissimilar character (8, 10, 11). For those whose primary interest is in the nature, development, and organization of psychological traits, such comparisons are a means of determining some of the crucial variables associated with the development and organization of particular kinds of trait patterns (1). For those who are especially interested in group differences, such comparisons permit a more comprehensive and analytical approach to the patterning and interrelationships among significant variables (2, 5, 12). And for those who are concerned about the appropriateness of certain types of tests for certain kinds of populations, such comparisons furnish data which can be of invaluable assistance in making the correct choice of tests. The findings of the present study may be interpreted in the light of any one of these considerations.

The purpose of this study was to compare the factorial organization of mental abilities for a well-defined High and Low Status Group. This necessitated the selection of a test battery that adequately represented the domain of mental abilities as charted by previous factorial experiments, the administration of this battery to a carefully selected High Status and Low Status Group, the computation of intercorrelation coefficients among the tests for

¹ This study is based on a Ph. D. dissertation done under the direction of Dr. Allison Davis of the University of Chicago.

each of the two groups, and the factorization of both correlation matrices with subsequent rotation of reference axes. The factor matrices so obtained were to be subject to a comparative analysis which would consider all evidence of significant similarities or differences in factor patterns but which would give special consideration to the following questions:

(1) What proportion of the common factor variance for both groups is accounted for by a general first-order factor corresponding to Spearman's general intellective factor "g"?

(2) Are essentially the same factors identified in the analyses for the two status groups?

(3) Is there any difference in the extent to which the "simple structure" criterion can be attained in applying rotational procedures to the two groups?

(4) What comparisons can be made between the organization of mental abilities for each group as determined by the correlations among the separate factors or abilities?

(5) Are the factorial compositions of individual tests in the battery approximately the same for the two groups, or are there differences that suggest that some of the tests represent different functions or organizations of such functions for the two groups?

PROCEDURE

The subjects of this study were eleven- and twelve-year-old pupils in certain selected schools located in and around the city of Chicago. Only those schools were selected whose student populations seemed likely to contain a large proportion of either "Low" or "High" status people as defined by the index of status employed in this study. All of the eleven- and twelve-year-old pupils in these selected schools were administered a battery of mental tests consisting of the following well-known and widely used instruments: the Addition, Multiplication, Vocabulary, Completion, Figures, Cards, First Letters, Four-Letter Words, Suffixes, Letter Series, Letter Grouping, and Pedigrees tests of the *Chicago Tests of Primary Mental Abilities* for ages eleven to seventeen; the Spatial Relationships, Logical Reasoning, Numerical Reasoning, and Verbal Concepts subtests which comprise the elementary form of the *California Short-Form Test of Mental Maturity*, but which were treated as separate tests for purposes of analysis; the *Otis Quick-Scoring Mental Ability Test* (Beta, Form Dm); and the *Henmon-Nelson Test of Mental Ability* (Elementary, Form A). The testing was done in four sessions to avoid the effects of fatigue.

The tests from the Chicago series are instruments of demonstrated factorial purity, and they were included in the present battery so that any evidence of differences in mental organization for the two status groups would be clearly revealed in the resulting configurations. The four subtests of the California test, which are also supposed to be representative of separate abilities or factors, were included for essentially the same reason, and it was judged that these sixteen tests would bear the major burden of defining and allocating the common-factor space for each of the status groups. The Otis Beta and Henmon-Nelson tests were included in the battery so that any differences in the rôle of the so-called "IQ" test in the mental organization of the separate status groups would be clearly revealed in the analysis.

At the end of the first testing session each subject received a questionnaire which he was requested to take home and have his parents complete. The most important function of this questionnaire was to determine the father's occupation and the address of the family dwelling, both of which were necessary to determine the index of status that was employed in this study. This index was a modified form of the "Index of Status Characteristics" developed by Warner, Meeker and Eells, which included scales for rating the father's occupation, the house type, the dwelling area and the source of income (15). The latter variable was omitted from the index used for the present study, because of anticipated resistance to questions concerning source of income, and the three remaining variables were combined to form a modified Index in accordance with a method suggested by Warner *et al.* for computing the Index when this variable is unavailable. The father's occupation was obtained directly from the questionnaires, and the type of house lived in and the nature of the dwelling area were easily ascertained by actually visiting the location after obtaining the address from the questionnaire and then observing and rating while on the scene. Since each child had an "ISC," or Index of Status Characteristics, which indicated his relative position on the status continuum, "Low" and "High" status groups could be defined by establishing certain critical ISC scores as limits for each group. Accordingly, the High Status Group was designated as occupying that part of the status continuum extending from an Index value of 33 or less, while the Low Status Group was designated as occupying that part of the continuum extending

from an Index value of 54 or more. (Low Index values are indicative of high status.) For the original population on which the ISC was validated, these critical values designated subjects who were upper middle class or above in the one case and upper lower class or below in the other (15). The ISC averages for the Low and High Status groups were 67.6 and 22.1 respectively, and the status range extended from upper-middle status to upper-upper status for the High Status Group and from lower-lower status to upper-lower status for the Low Status Group. There were one hundred and seventy-nine subjects in the High Status Group and one hundred and seventy-two subjects in the Low Status Group thus defined. "Status," as employed herein and measured by the ISC, actually refers to socio-economic status, but the correlation between socio-economic level and social status is high enough (0.966) to justify using these terms interchangeably for our present purposes (15).

In preparing the data for analysis, coefficients of correlation were computed for all possible test pairs within each of the status groups, and these correlations were recorded in the 18 x 18 correlation matrices that were prepared separately for both the Low and High Status Groups.² Both of these matrices were factored by means of the centroid method of multiple factor analysis as described by L. L. Thurstone, and the factor matrices so obtained were subject to rotational procedures until oblique solutions were achieved which best satisfied the criteria for simple structure (14). The resulting factor patterns for the two groups were then compared.

Because the correlations between primary factors are determined by the placement of the reference axes in accordance with simple structure criteria, which is at best a somewhat inexact procedure, those correlations tend to be rather unstable. Therefore, in order to obtain reliable data bearing on the correlational relationships between the factors that were extracted, multiple group analyses of the correlation matrices for both status groups were undertaken, employing for this purpose the factorial clusters that had been identified in the centroid analysis (14). This proce-

² The original correlation matrices have been deposited with the American Documentation Institute. Order Document No. 4866 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Remit \$1.25 for 35 mm microfilm or \$1.25 for photoprints.

ture yielded correlations between factors that were unique and mathematically determinate and which could be used confidently for purposes of comparison.

RESULTS AND INTERPRETATION

The centroid factor matrices that resulted from the factorization of the correlation matrices for the two status groups are shown in Table I (for the Low Status Group) and Table II (for the High Status Group). Factor extraction was regarded as complete when two-thirds of the loadings for a factor fell below twice their standard errors (as determined by Burt's empirical formula) (13), and when the product of the two highest loadings for a factor fell below $\frac{1}{\sqrt{N}}$ (13). Application of these combined criteria indicated that factorization should be concluded for both groups with the extraction of the sixth factor, but a seventh factor was also extracted to make absolutely certain that further meaningful factors

TABLE I—CENTROID FACTOR MATRIX: LOW STATUS GROUP

Test	Factor							
	I	II	III	IV	V	VI	VII	h ²
PMA Addition	59	28	-32	-24	28	13	-15	70
PMA Multiplication	63	32	-27	-22	24	14	-11	71
PMA Vocabulary	76	15	17	16	-03	-20	-15	72
PMA Completion	78	03	17	-06	-07	-15	04	67
PMA Figures	48	-42	-27	21	-19	05	-15	58
PMA Cards	58	-44	-37	19	-28	03	05	78
PMA First Letters	74	29	10	24	-05	23	11	77
PMA Four-Letter Words	59	37	-13	-09	-28	05	16	62
PMA Suffixes	61	29	17	07	-30	24	-10	65
PMA Letter Series	64	-20	12	23	40	07	19	72
PMA Letter Grouping	53	09	-14	24	32	-11	23	53
PMA Pedigrees	66	-04	10	26	16	-04	-07	55
Calif. Spatial Relations	56	-19	-14	-03	-04	-05	12	39
Calif. Logical Reasoning	71	-22	09	-19	-01	-04	-07	60
Calif. Numerical Reasoning	74	-15	05	-20	13	11	09	65
Calif. Verbal Concepts	84	08	28	-04	-06	-08	05	80
Otis Beta	89	-09	24	-24	-02	03	03	92
Henmon-Nelson	87	-15	20	-21	-07	16	10	90
Per cent of Total Variance	47	06	04	04	04	02	02	

TABLE II—CENTROID FACTOR MATRIX: HIGH STATUS GROUP

Test	Factor							
	I	II	III	IV	V	VI	VII	h^2
PMA Addition	61	32	-29	-22	33	12	11	74
PMA Multiplication	56	40	-29	-23	37	22	13	81
PMA Vocabulary	73	22	45	17	-10	07	10	84
PMA Completion	76	07	27	30	-04	-11	15	78
PMA Figures	44	-26	-50	28	-22	-03	23	69
PMA Cards	52	-43	-41	25	-19	10	16	76
PMA First Letters	50	42	-15	-04	-23	-12	-07	52
PMA Four-Letter Words	42	30	-20	-12	-27	-24	-16	48
PMA Suffixes	55	33	-14	03	-13	12	-30	55
PMA Letter Series	64	-20	15	-28	-24	13	-10	64
PMA Letter Grouping	49	-16	06	-37	-15	06	09	44
PMA Pedigrees	74	-18	25	-21	-05	11	-09	71
Calif. Spatial Relations	56	-27	-12	14	14	07	-21	49
Calif. Logical Reasoning	54	-33	18	06	24	-13	-15	53
Calif. Numerical Reasoning	70	-29	-04	-04	21	-14	08	65
Calif. Verbal Concepts	82	20	35	16	05	09	11	88
Otis Beta	86	-09	22	10	13	-10	04	83
Henmon-Nelson	80	-05	24	12	13	-06	-11	75
Per cent of Total Variance	41	08	07	04	04	02	02	

could not be identified. After the rotation of reference axes, it was found that the seventh factor was a residual factor of no importance, and it was therefore discarded for both groups. The first six factors represented sixty-seven per cent of the total test variance for the Low Status Group and sixty-six per cent of the total test variance for the High Status Group.

Inspection of Tables I and II reveals an interesting difference in first factor variance for the two groups. This first unrotated centroid factor, which roughly corresponds to Spearman's general intellectual factor "g," accounts for some forty-seven per cent of the total test variance for the Low Status Group and only forty-one per cent of the total test variance for the High Status Group. This amounts to about seventy per cent of the common-factor variance for the Low Status Group and sixty-two per cent of the common-factor variance for the High Status Group. This difference of six per cent for the total variance and eight per cent for the common-factor variance indicates that for the Low Status

TABLE III—FINAL FACTOR MATRIX: OBLIQUE
SOLUTION FOR LOW STATUS GROUP

Test	Factor					
	V	WF	S	R	N	GT
PMA Addition	00	-.04	01	00	52	00
PMA Multiplication	00	01	-.01	00	48	01
PMA Vocabulary	40	02	06	05	11	-.14
PMA Completion	23	-.06	-.01	-.07	-.05	17
PMA Figures	06	01	62	03	09	-.07
PMA Cards	-.04	-.01	66	-.07	-.05	04
PMA First Letters	-.07	42	-.02	24	02	-.07
PMA Four-Letter Words	00	18	04	-.25	-.03	10
PMA Suffixes	07	43	06	00	05	-.02
PMA Letter Series	-.12	03	-.06	50	-.01	06
PMA Letter Grouping	-.02	-.05	-.03	27	06	-.11
PMA Pedigrees	17	06	08	29	11	-.13
Calif. Spatial Relations	00	-.11	21	-.05	-.04	17
Calif. Logical Reasoning	15	-.15	12	-.05	04	29
Calif. Numerical Reasoning	-.07	-.06	00	08	04	35
Calif. Verbal Concepts	21	04	-.09	01	-.07	19
Otis Beta	10	-.04	-.04	-.03	-.03	39
Henmon-Nelson	-.06	06	00	01	-.09	44

Group there is more of a communality of function between the various tests that make up the battery, and, if the present test battery adequately represents the domain of mental abilities, it is strong evidence for a less differentiated organization of mental abilities for the Low Status Group. Apparently the Low Status child approaches his various mental activities with a relatively unspecialized intellectual structure, and his "general" intelligence (or "g") is in fact a relatively more generalized kind of function than its High Status counterpart and plays a correspondingly greater rôle in the solution of all of the more specialized types of intellectual problems as well as those of a general nature. These observations are also confirmed by other data that will be presented.

Rotation of reference axes in accordance with simple structure criteria resulted in the oblique solutions shown in Table III for the Low Status Group and Table IV for the High Status Group. The oblique solution was considered preferable for both groups, since inspection of the correlation matrices disclosed that the

TABLE IV—FINAL FACTOR MATRIX: OBLIQUE
SOLUTION FOR HIGH STATUS GROUP

Test	Factor					
	V	WF	S	R	N	GT
PMA Addition	-.02	.03	.02	.00	.61	.07
PMA Multiplication	.01	-.05	-.01	.00	.69	-.01
PMA Vocabulary	.62	.00	-.05	.08	-.03	-.09
PMA Completion	.55	.10	.12	-.12	-.05	.12
PMA Figures	.01	.07	.68	-.08	.03	.02
PMA Cards	-.02	-.09	.64	.05	-.03	-.01
PMA First Letters	.11	.36	.02	-.01	.08	-.06
PMA Four-Letter Words	-.07	.46	-.01	.01	-.02	.05
PMA Suffixes	.05	.07	-.10	.03	.00	-.15
PMA Letter Series	-.03	.02	.01	.52	-.08	-.06
PMA Letter Grouping	-.07	.08	.09	.47	.11	.00
PMA Pedigrees	.08	-.04	-.08	.40	-.02	.06
Calif. Spatial Relations	-.07	-.14	.06	-.02	-.07	.18
Calif. Logical Reasoning	.02	-.03	-.11	-.02	-.11	.40
Calif. Numerical Reasoning	.00	.07	.14	.03	.14	.39
Calif. Verbal Concepts	.57	-.04	-.03	.03	.10	-.01
Otis Beta	.31	.06	.02	.00	.03	.28
Henmon-Nelson	.27	.02	-.10	-.01	-.05	.24

inter-factor correlations would probably be too high to justify the imposition of orthogonal structure on the test configurations. Essentially the same factors were identified for both groups, and these factors, with but one exception, represented the primary mental abilities that have been reported by Thurstone and others: The "V" or Verbal Meaning factor, which represents a knowledge of words and their meanings, had its highest loadings for the Vocabulary and Completion tests of the Chicago PMA series and the Verbal Concepts test of the California for both status groups. The "WF" or Word Fluency factor, which represents the ability to call quickly to mind the word forms which are appropriate or required for given circumstances or conditions, had its highest loadings for the First Letters, Four-Letter Words, and Suffixes tests for the Low Status Group, while the same factor was determined primarily by the First Letters and Four-Letter Words tests for the High Status Group. The "S" or Space factor, which involves the comprehension of spatial relations or the ability to visualize the manipulation of objects in space, was determined

primarily by the Figures and Cards tests of the Chicago PMA series for both status groups. The "R" or Reasoning factor, which refers to the skill required to discover a principle or relationship from a given set of data and then to make a further application of it, was determined primarily by the Letter Series, Letter Grouping, and Pedigrees tests of the Chicago PMA series for both status groups. The "N" or Number factor, which represents the ability to perform the simple numerical operations in addition, subtraction, etc., had its highest loadings for the Addition and Multiplication tests of the Chicago PMA series for both status groups. Also extracted from the correlation matrices for both status groups was a factor that had its highest loadings for the Otis Beta and Henmon-Nelson tests and the subtests for the California Test of Mental Maturity. Since the Otis Beta and Henmon-Nelson tests are both tests of general intelligence characterized by a high degree of factorial complexity, it is probably best to designate this factor as the "GT" or General Test factor. It is noteworthy that the subtests of the California, which are supposed to be measures of many of the same factors that were identified in the present analysis and which were originally expected to assist in the determination of those factors, actually contributed very little to their determination. Only the Verbal Concepts test was useful in this respect, while the others tended to cluster with the general ability tests and to have loadings on the General Test factor. The factorial impurity of the Spatial Relations, Logical Reasoning, and Numerical Reasoning subtests of the California may be due to the fact that each of these subtests actually requires the testee to perform two somewhat different kinds of tasks, both of which are supposed to be satisfactory and highly correlated measures of the ability in question. That the combination of the scores for these separate tasks into a so-called "factor" score seems to result in an index that is more closely related to general mental ability than to any of the primary mental abilities suggests that the two separate tasks which are involved in the "factor" subtests are either not highly correlated or are too general to be valid measures of a primary ability.

Even though the factors that were extracted from the correlation matrices for the two status groups were essentially the same, the factorial structure which was obtained after the rotation of the reference axes was much more clear-cut and definitive for the

High Status Group than for the Low Status Group. This is indicated by the fact that the tests that determined a given factor generally had higher loadings on that factor for the High Status Group and yet had projections on other axes that were distributed much more closely around the zero point of these axes. The oblique solution for the High Status Group seemed to meet the requirements of "simple structure" more adequately than that for the Low Status Group, and the final placement of the reference axes was somewhat more indeterminate for the Low Status Group. For the High Status Group this apparently betokens a state of affairs in which the tests tend to cluster closely together to determine specialized ability groupings that differentiate the total test space into rather clear-cut common factor spaces. For the Low Status Group this differentiation does not exist in nearly as definitive a fashion. This is observed most clearly when the various factors are plotted against one another by graphical methods. The lack of differentiation for the Low Status Group is then evident for all factors, but is particularly evident in the plots of V against R, WF against R, WF against V, V against GT, and GT against R.

The Verbal Meaning factor for the Low Status Group is a good example of a factor that is very poorly defined in comparison with its High Status counterpart. The tests that determine this factor for both status groups are the Vocabulary and Completion tests of the Chicago series and the Verbal Concepts test of the California, but the loadings for these tests are 0.62, 0.55, and 0.57, respectively, for the High Status Group and 0.40, 0.23, and 0.21, respectively, for the Low Status Group. This is rather unusual in view of the important rôle that verbal ability is supposed to take in determining mental test performance for the Low Status Group. A reasonable explanation for this rather appreciable difference in factor loadings is that verbal skill is an extremely generalized function for the Low Status Group and thus permeates all mental tasks to such a degree that it loses its factorial identity in the first order and becomes almost collinear with a general ability factor of the second order.

Evidence in favor of this interpretation is found in Tables V and VI, which show the correlations between group centroid axes which resulted when multiple group analyses were undertaken on the correlation matrices for both status groups. The same tables also show the projection of each axis on the second-order general

TABLE V—CORRELATIONS BETWEEN GROUP CENTROID AXES AND LOADINGS ON SECOND-ORDER GENERAL FACTOR: LOW STATUS GROUP

Factor	Factor					
	V	WF	S	R	N	GT
V (Verbal Meaning)						
WF (Word Fluency)	79					
S (Space)	58	45				
R (Reasoning)	77	61	60			
N (Number)	72	68	51	68		
GT (General Test)	88	72	67	71	77	
r_{mg} (2nd-Order Loadings)	93	78	65	82	81	94

factor. The tests that were chosen for the various groups followed rather closely the factorial structure revealed in the centroid analyses, but an effort was made to include at least three tests in each group. The Number factor was represented by the Addition and Multiplication tests of the Chicago series and the Numerical Reasoning test of the California. The Verbal Meaning factor was represented by the Vocabulary and Completion tests of the Chicago series and the Verbal Concepts test of the California. The Space factor was represented by the Figures and Cards tests of the Chicago series and the Spatial Relations test of the California. The Word Fluency factor was represented by the First Letters, Four-Letter Words, and Suffixes tests of the Chicago series. The

TABLE VI—CORRELATIONS BETWEEN GROUP CENTROID AXES AND LOADINGS ON SECOND-ORDER GENERAL FACTOR: HIGH STATUS GROUP

Factor	Factor					
	V	WF	S	R	N	GT
V (Verbal Meaning)						
WF (Word Fluency)	58					
S (Space)	41	37				
R (Reasoning)	67	50	51			
N (Number)	59	60	52	60		
GT (General Test)	85	45	58	76	67	
r_{mg} (2nd-Order Loadings)	83	63	60	81	79	90

Reasoning factor was represented by the Letter Series, Letter Grouping, and Pedigrees tests of the Chicago series. And the General Test factor was represented by the Logical Reasoning test of the California and the Otis Beta and Henmon-Nelson general intelligence tests. These same groups were employed in the multiple group analyses for both status groups. The factor matrices which were obtained through this procedure did not yield any additional data beyond that already obtained from the centroid analyses, but the correlations between the group centroid axes which were secured as a by-product of this method revealed some very interesting relationships.

In Tables V and VI it can be observed that all the correlations except two are higher for the Low Status Group, and these discrepancies are particularly large for the Verbal Meaning and Word Fluency factors. In fact, the correlations of the Verbal Meaning factor with the other factors are so high for the Low Status Group that this factor has a loading of 0.93 on the second-order general intellective factor, making the one almost collinear with the other. This loading is about as high as that for the General Test factor, establishing rather convincingly the importance of verbal facility in the determination of tested intelligence for the Low Status Group. All of the other factors at the Low Status level also have higher loadings on the second-order general factor than their High Status counterparts. Apparently each of the primary mental abilities is more a function of a second-order parameter for the Low Status Group than for the High Status Group, and this second-order parameter is more thoroughly saturated with verbal components for that group. This suggests that a certain minimum level of verbal competence must be attained before a test of any special ability can become a more specific measure of that primary function, and for many Low Status youngsters this level has not been attained. All tests, both general intelligence tests and primary ability tests, are then more heavily weighted with the verbal component for the Low Status Group and would to this extent be less valid as measures of the particular ability or combination of abilities they purport to measure. This would be true, of course, of all tests except those specifically designed as tests of verbal competency.

This tendency for the primary mental functions of the Low Status Group to cluster around a general-verbal axis of the second

order is again indicative of the relative lack of differentiation that exists in the organization of mental abilities for the Low Status Group. This lack of differentiation has been manifest in three ways: in terms of the lack of definitiveness of the factors extracted, in terms of the amount of variance accounted for by the first centroid, and now in terms of the relative lack of independence among the primary abilities and the concomitant tendency for these abilities to cluster around a general-verbal factor of the second order.

Other interesting comparisons can be made from the data presented in Tables V and VI. Particularly interesting are the relationships between the Word Fluency factor and the other factors for the two groups. Like the Verbal Meaning factor, the Word Fluency factor has a much higher loading on the second-order general intellectual factor for the Low Status Group, showing an increase from 0.63 to 0.78. This increase in general factor loading is similar to conditions in the first order, where the three tests that determine the Word Fluency factor—First Letters, Four-Letter Words, and Suffixes—can be observed to have considerably higher loadings on the first centroid for the Low Status Group (Tables I and II). This same Word Fluency factor is also much more highly correlated with the General Test and Verbal factors for the Low Status Group, increasing from 0.45 to 0.72 for the former and from 0.58 to 0.79 for the latter. Word fluency would appear to be a relatively independent ability at the High Status level, but at the Low Status level it is much less differentiated from the verbal-general ability complex as defined by the first and second order general factors of this study. This lack of differentiation is accompanied by an interesting change in factorial character. For the Low Status Group the Word Fluency factor is most highly correlated with the Verbal Meaning factor and then with the General Test and Number factors. For the High Status Group its highest correlation is with the Number factor. Again it would seem that a minimum level of verbal competency is required before we can presume to measure the special abilities, for a child cannot be fluent until he has a considerable repertoire of words with which to be fluent, and after that his fluency is limited only by his speed of association, which is a requisite for success in number computation and word fluency alike.

The comparisons which have been made between the two status

groups were those which seemed most significant to the author, but there are other differences which might also be considered. Among these are differences in individual test loadings for the two groups. It might be asked why the Suffices test helped to determine the Word Fluency factor for the Low Status Group but not for the High Status Group, or why the Logical Reasoning test of the California had a much higher loading on the first centroid for the Low Status Group, or why the Pedigrees test of the Chicago series had a much higher loading on the first centroid for the High Status Group. It is beyond the scope of the present paper to consider these minor differences in detail, but the reader is invited to interpret these differences in the light of any explanations that appear to be logical and justified in terms of the data presented.

It might be contended that the relatively undifferentiated organization of mental abilities found for the Low Status Group in this study was actually an artifact of either or both of two irrelevant conditions: (a) a greater range of ability for the Low Status Group, resulting in generally higher variances and correspondingly higher correlations for that group; or (b) the influence of general performance level on the structure of abilities, which, because it would be lower for the Low Status Group, would leave in question whether this lack of differentiation was associated with Low Status *per se* or merely with the lower level of general performance associated with the Low Status Group. Evidence against the former interpretation emerged from a comparison made between the standard deviations computed for both groups for all tests, in which three standard deviations were observed to be significantly higher for the Low Status Group compared to two that were significantly higher for the High Status Group. The latter two were both for tests in the Chicago series, which were the most instrumental in establishing the evidence for a lack of differentiation among mental abilities for the Low Status Group. Evidence against the latter interpretation arose as a result of comparisons that were made between the factor matrices of a Low and a High Status Group which were "matched" for Otis Beta IQ but were still comparable to the original status groups in terms of their average indices of social status. Although this procedure is not without disadvantages, the evidence from this comparison again seemed to indicate a greater differentiation of mental abilities for the

High Status Group and a corresponding lack of differentiation for the Low Status Group.

This differentiation of traits, which has been reported as lacking for the mental abilities of the Low Status Group, is certainly not a new concept in the field of developmental psychology. As a theoretical construct it has served to organize and give meaning to data in the areas of mental, physical, and emotional growth as well as to the growth process as a whole (3, 4, 7, 9). It has made one of its most recent appearances in the form of a hypothesis purporting to be a developmental principle of mental growth. Proposed by H. E. Garrett, this hypothesis states that: "Abstract or symbol intelligence changes in its organization as age increases from a fairly unified and general ability to a loosely organized group of abilities or factors" (6).

Although there is somewhat conflicting evidence about the differentiation of abilities with age, most studies in this area seem to report trends in the direction stated by the hypothesis. Assuming for the moment that this hypothesis has been adequately confirmed, we can combine this principle with the results of the present study and conjecture that the differentiation of mental abilities is not only a function of increasing age but also a function of the quality of the environment and the type of attendant mental stimulation which constantly acts upon the individual during the growth process. Optimum mental stimulation would facilitate the maximum differentiation of mental abilities. Since Low Status youngsters are far less subject to the type of intellectual stimulation that is conducive to the development of the abstract, academic skills that are so highly valued by the middle class, they would therefore possess a relatively undifferentiated organization of mental abilities. Additional data in the form of comparative studies of the structure of abilities for other groups would help to clarify the nature of the specific characteristics within the environment that are the most important agents in the facilitation or inhibition of this process of differentiation.

SUMMARY

The purpose of this study was to compare the factorial organization of mental abilities for a well-defined High and Low Social Status Group. To accomplish this purpose a comprehensive bat-

tery of mental tests was administered to a large group of eleven- and twelve-year-old subjects. This battery consisted of the Addition, Multiplication, Vocabulary, Completion, Figures, Cards, First Letters, Four-Letter Words, Suffixes, Letter Series, Letter Grouping, and Pedigrees tests of the *Chicago Tests of Primary Mental Abilities* for ages eleven to seventeen; the Spatial Relationships, Logical Reasoning, Numerical Reasoning, and Verbal Concepts subtests of the *California Short-Form Test of Mental Maturity*; the *Otis Beta Quick-Scoring Mental Ability Test*; and the elementary form of the *Henmon-Nelson Test of Mental Ability*. The subjects were classified into the High or Low Status Groups according to their scores on the "Index of Status Characteristics," and the final analysis was made with one hundred and seventy-two subjects in the Low Status Group and one hundred and seventy-nine subjects in the High Status Group (15). Scores on the eighteen tests in the battery were intercorrelated for each group, and the correlation matrices so obtained were factored by means of Thurstone's complete centroid method (14). The resulting factor patterns for the two groups were then compared. A Multiple Group analysis was also undertaken in order to compare the correlations between the factors extracted.

Essentially the same factors were identified for both status groups: Number, Verbal Meaning, Space, Word Fluency, Reasoning, and a General Test factor for which all the general ability tests had high loadings. But although the factors were essentially the same, the organization of mental abilities was found to be much less differentiated for the Low Status Group. This lack of differentiation in the organization of mental abilities for the Low Status Group was evident in three principal ways:

- (1) The first centroid factor, similar to Spearman's "g" and expressing the communality of function between the several variables employed, accounted for a greater proportion of common factor variance for the Low Status Group than for the High Status Group.

- (2) The factors extracted for the Low Status Group were not nearly as clear-cut and definitive as those for the High Status Group, and there was a sharpness and clarity in the final "simple structure" solution for the High Status Group that was entirely lacking for the Low Status Group.

(3) There was a relative lack of statistical independence among the factors extracted for the Low Status Group (i.e., the factors were more highly intercorrelated), and there were correspondingly higher loadings for these factors on the second-order general factor.

The rôles played by the Verbal Meaning and Word Fluency factors in the factor patterns for the two status groups suggested that the lack of differentiation associated with the Low Status Group was at least partly the result of the increased saturation of the general factor with verbal components. This condition apparently arises as a result of the fact that all mental tests, including the tests of primary or special abilities, become more verbally weighted for the Low Status youngster because he has not attained the requisite level of verbal competence to execute easily the modest verbal tasks that are inevitably involved in even the most "nonverbal" of tests.

REFERENCES

- 1) A. Anastasi and J. P. Foley, *Differential Psychology*, pp. 512-520. New York, Macmillan, 1949.
- 2) B. Balinsky, "An Analysis of the Mental Factors of Various Age Groups from Nine to Sixty," *Genetic Psychology Monographs*, 23: 191-234, 1941.
- 3) K. M. B. Bridges, "Emotional Development in Early Infancy," *Child Development* 3: 324-334, 1932.
- 4) G. E. Coghill, "Individuation Versus Integration in the Development of Behavior," *Journal of General Psychology*, 3: 431-435, 1930.
- 5) H. E. Garrett, A. I. Bryan, and R. E. Perl, "The Age Factor in Mental Organization," *Archives of Psychology*, No. 176, 1935.
- 6) H. E. Garrett, "A Developmental Theory of Intelligence," *American Psychologist*, 1: 372-378, 1946.
- 7) A. Gesell and H. Thompson, *Infant Behavior: Its Genesis and Growth*. New York, McGraw-Hill, 1934.
- 8) K. J. Holzinger and F. Swineford, "A Study in Factor Analysis: the Stability of a Bi-factor Solution," *Supplementary Educational Monographs*, No. 48. Chicago, University of Chicago Press, 1939.
- 9) M. Sherman, "The Differentiation of Emotional Responses in Infants," *Journal of Comparative Psychology*, VII: 265-284, 335-351, 1927.
- 10) F. Swineford, "Growth in the General and Verbal Bi-factors from Grade VII to IX," *Journal of Educational Psychology*, 38: 257-272, 1947.
- 11) F. Swineford, "General, Verbal and Spatial Bi-factors after Three Years," *Journal of Educational Psychology*, 40: 353-360, 1949.
- 12) W. H. Taylor and J. G. Darley, "Intercorrelations of the Test Scores Made by Selected Samples of Unemployed Men and Women," *Univ. Minn., Empl. Stab. Res. Inst. Bull.*, 3: 203-207, 1934.

- 13) G. Thomson, *The Factorial Analysis of Human Ability*, pp. 122-124. London, University of London Press, 1951.
- 14) L. L. Thurstone, *Multiple Factor Analysis*, pp. 149-176. Chicago, University of Chicago Press, 1947.
- 15) W. L. Warner, M. Meeker, and K. Eells, *Social Class in America*. Chicago, Science Research Associates, 1949.

EFFECT OF IMMEDIACY OF KNOWLEDGE OF CORRECTNESS OF RESPONSE UPON LEARNING

ARTHUR C. F. GILBERT

University of Nebraska

One of the existing problems in education is that of making the testing period a profitable learning experience for students. Some instructors have followed the practice of discussing the test items with the students as soon as possible after the completion of an examination. Others have provided the students with "model" responses similar to those involved in an examination. Still other instructors have given their attention to developing devices which permit the student to determine the correctness of his response as soon as the responses have been made.

The devices provided by the latter group of experimenters have been many and ingenious. Among such devices is the Science Research Associates (S.R.A.) Self-Scorer invented by Angell and Troyer (2). This device was made available commercially in 1949 and is designed for use with objective tests of either the multiple choice, matching or true-false variety.

The S.R.A. Self-Scorer consists of four parts, a combination answer sheet and key, a perforated mat, and a plastic frame or envelope. In the assembly of the Self-Scorer, the perforated mat is inserted between the answer-sheet and key, and this assembly is then inserted in the plastic envelope. In using the Self-Scorer the student punches a spot on his answer sheet corresponding to what he believes to be the correct response. If his selection of responses is correct a red dot appears on his answer sheet, but if his selection of response is incorrect, the answer sheet remains white at that place.

A review of the literature indicates that research using self-scoring devices has been going on for at least thirty years. Pressey (7) in 1926 reported the development of a self-scoring device which kept the problem before the student until he responded correctly. Among the research aimed at increased efficiency in learning through immediacy of knowledge of correctness of response are studies by Little (5), Peterson (6) and Stephens (8). Studies

involving the use of the Self-Scorer were made by Angell (1), Jones and Sawyer (4), Fields (9) and Blood (3). A consideration of previous studies in this area has indicated the need for a study which would test certain hypotheses within the framework of one research design while using the same group of subjects.

The purpose of this study, therefore, was to investigate the following three implications: (1) the effect of the S.R.A. Self-Scorer compared with the traditional method of testing followed by class discussion of the test items, when gain from two administrations of the same test is used as a criterion and with scholastic aptitude and achievement scores obtained from a preregistration test battery held constant; (2) the effectiveness of the S.R.A. Self-Scorer evaluated under the same conditions but with the score on a separate test as the criterion; (3) the differential effectiveness of the two methods of testing as regards learning outcomes when the test items are classified according to the educational outcomes which they are designed to measure.

PROCEDURE

Students of the beginning course in educational psychology at the University of Nebraska were used as subjects for the investigation. Subjects from eight sections, four instructors teaching two sections each, were used. One of the two sections taught by each instructor was designated as belonging to the experimental group, and the other section was used for the control group. In each of the experimental and the control groups there were four sections. There were ninety-eight students in this experimental group and one hundred and fourteen in the control group.

At the end of the first six weeks of the semester an examination covering the instruction up to that point was administered. This test was composed of seventy-four objective type items of the multiple choice and matching variety. The experimental group used the Self-Scorer, whereas the control group used standard type answer sheets.

The six-weeks examination was administered to the experimental and the control group in two different forms. Form A covered the entire six weeks of instruction and was administered to the control group during a fifty-minute class period. Form B contained the identical items of Form A but was divided into two parts; Part I

contained the first thirty-six items of Form A, and Part II contained the subsequent thirty-eight items.

The control group was allowed one fifty-minute period for the completion of the total test. The next class period was utilized for discussion of the test items. At this time the test booklet (form A) was returned to the students, together with their answer sheets. The instructors advised the students that notes were not to be taken during this period. After discussion of the test items all test forms and answer sheets were returned to the instructors.

The first part of the six-weeks test (Form B) was administered to the experimental classes during the first class period; the second class period was used for the remainder of the test by the experimental group; there was no discussion of any kind with this group.

Prior to the administration of the test to the experimental group the nature of the Self-Scorer was explained. Students were advised that the device was so constructed as to permit them to know whether they were responding correctly to the items while taking the test. The procedure to be followed while taking the test was explained.

Item analyses were made of all items in this first test. The total number of persons responding incorrectly to each item was tabulated and a post-test was constructed involving those concepts measured by the items most frequently missed in the original test. Four weeks later the original test was again administered to both the experimental and the control groups, together with the twenty-five-item post-test. Standard answer sheets were used by both groups for the second testing period. Both groups were tested without any warning on the same day.

The data from this investigation were analyzed by the analysis of covariance. Two criteria were employed, the first criterion being the gain on the six-weeks examination and the second criterion the score on the post-test. The control variables used were the scores on a locally constructed preregistration English Placement examination, the scores on the Linguistic subtest and the scores on the Quantitative subtest of the American Council on Education (ACE) Psychological Examination.

Items of the test were classified according to type of behavior measured in terms of course objectives and responses to each item tabulated for both administrations of the same test for both the

experimental and the control group. These data were used to determine the effect of the Self-Scorer on learning when different types of learning outcomes are considered.

RESULTS

Gain as a criterion. The first criterion used was the gain in scores on the second administration of the same test. The analysis of covariance was used to test the significance of the difference between the means of the two groups holding the control variables constant. In Table I are shown the results of the analysis of covariance. The value of F of 13.93 with 1 and 207 degrees of freedom is significant at the one per cent level of confidence.

By comparison of the two adjusted criteria means it may be seen that the students in the control group are favored over those in the experimental group. Therefore, it may be concluded that, insofar as no other pertinent factors contribute a bias, the evidence indicates that the gain achieved by the students in the control group

TABLE I—SIGNIFICANCE OF THE INFLUENCE OF THE S.R.A. SELF-SCORER ON THE GAIN ON THE SAME TEST WITH CONTROLS (ANALYSIS OF COVARIANCE)

Source of Variation	Degrees of Freedom	Residuals	
		Sum of squares	Mean square
Total			
Within subgroups	208	7115.50	
	207	6666.88	32.2071
Difference	1	448.62	448.62

$$F_{1,207} = \frac{448.62}{32.2071} = 13.93$$

$$t = \sqrt{F} = \sqrt{13.93} = 3.73$$

Adjustments:

$$\begin{aligned}\bar{Y}_B &= (\bar{X}_{1B} - \bar{X}_{2T})a_1 + (\bar{X}_{2B} - \bar{X}_{2T})a_2 + (\bar{X}_{3B} - \bar{X}_{3T})a_3 \\ &= .37a_1 + .33a_2 + .12a_3 \\ &= .213\end{aligned}$$

$$\begin{aligned}\bar{Y}_C &= (\bar{X}_{1C} - \bar{X}_{1T})a_1 + (\bar{X}_{2C} - \bar{X}_{2T})a_2 + (\bar{X}_{3C} - \bar{X}_{3T})a_3 \\ &= -.32a_1 + (-.27)a_2 + (-.10)a_3 \\ &= -.178\end{aligned}$$

$$\begin{aligned}a_1 &= .26152626 \quad \bar{Y}_B = 1.79 - .21 = 1.58 \\ a_2 &= .29826304 \quad \bar{Y}_C = 4.29 + .18 = 4.47 \\ a_3 &= .14161260\end{aligned}$$

was significantly greater than the gain achieved by those students in the experimental group. Thus, the hypothesis "there is no significant difference between the two group means for the second administration of the same test with the English Placement scores, the L scores and the Q scores controlled," is rejected.

The significance of the relationship between the criterion and the control variables in this investigation was tested by the analysis of regression. The obtained F value of 1.11 with 3 and 208 degrees of freedom is not significant. It may also be concluded that the obtained multiple correlation coefficient $Ry(x_1x_2x_3)$ of 0.126 is not significantly different from zero. Thus, it may be concluded that evidence is provided that the control variables did not account for a significant amount of variation in the criterion beyond that accounted for by the experimental-control group classification. To the extent to which validity is characteristic of the measurements employed, evidence from the analysis of results with gain as a criterion indicates that the students using the traditional answer sheets in the testing situation followed by class discussion at a later date achieve significantly higher when the test is readministered than students using the Self-Scorer without the benefit of discussion.

Post-test as a criterion. The second criterion used in determining the influence of the Self-Scorer on learning was the score on a twenty-five-item test which will be referred to hereafter as the post-test. As in the analysis of the results with gain as a criterion the analysis of covariance technique was employed for those data.

TABLE II—SIGNIFICANCE OF THE INFLUENCE OF THE S.R.A. SELF-SCORER ON THE SCORES ON THE POST TEST WITH CONTROL (ANALYSIS OF COVARIANCE)

Source of Variation	Degrees of Freedom	Residuals	
		Sum of squares	Mean square
Total	208	1327.86	6.33
Within subgroups	207	1309.62	
Difference	1	18.24	18.24

$$F_{1,207} = \frac{18.24}{6.33} = 2.88$$

$$t = \sqrt{F} = \sqrt{2.88} = 1.70$$

In Table II are shown the results of the second analysis of covariance. The value of F of 2.88 with 1 and 207 degrees of freedom is not significant. Therefore, the hypothesis, "there is no significant difference between the two group means for the scores on the post-test with the English Placement scores, the L scores and the Q scores controlled," is not rejected. It has not been possible to demonstrate that other than a single statistical population is represented by both these groups.

The significance of the relationship between the post-test as a criterion and the control variables was tested by the analysis of regression. An F value of 6.90 with 3 and 208 degrees of freedom was obtained. This is a greater value than that required at the one per cent level. Therefore, it may be concluded that the multiple correlation coefficient $R_{y_1(x_1x_2x_3)}$ of 0.50 is significantly different from zero. Evidence of the effectiveness of the controls in eliminating individual differences in the characteristics used as controls in this analysis is provided by the rejection of the zero correlation hypothesis.

To the extent to which validity is characteristic of the measurements employed, evidence from the analyses of results with the post-test scores as the criterion indicates that the contribution of the use of the Self-Scorer is not significantly greater than the traditional approach for students similar to those participating in the experiment under the same conditions.

Influence on different types of learning. Test items in the seventy-four-item test were classified into three categories: (1) definition items which involved the definition of psychological terms presented in class, (2) items involving the knowledge of facts, and (3) items requiring application of principles. A comparison of the responses of both the experimental and the control groups on both administrations of the same test was made for the three types of items. The responses were tabulated to determine the percentage of subjects in each group who responded incorrectly to the item on the first administration of the test but correctly the second time.

Analysis of the responses to the different types of test items shows that the experimental group had an equal or higher percentage of successes on four out of ten, or 40.0 per cent of the items which involved the definition of terms. Of the items which involved the knowledge of facts the experimental group had an equal or higher percentage on ten out of twenty-six, or 58.5 per cent of the items.

On the items which involved application of principles the experimental group was equal or higher percentage-wise on fifteen out of thirty-eight, or 39.5 per cent of the items.

No significant difference was found between the paired percentages for each of the three types of items. From the results of this analysis evidence is provided that the S.R.A. Self-Scorer does not significantly influence learning on one type of item more than others according to the classification of items used in this study, definitions, knowledge of facts and application of principles.

DISCUSSION

In this study it was discovered that, in general, students actually required twice the length of time to complete a test by use of the S.R.A. Self-Scorer than by other methods. Therefore, from the viewpoint of administration the use of the S.R.A. Self-Scorer would have to be justified by the fact that those using this device could perform equally with those tested in the traditional method in which one period is used for testing and the next one being used for discussion.

The results of this study indicated that when gain on the same test was used as a criterion those using the standard answer sheets accompanied by discussion performed significantly better than those using the S.R.A. Self-Scorer. If the experimental group had more experience with the S.R.A. Self-Scorer, it is possible that they might have shown a higher performance. On the twenty-five-item test there was no significant difference between the two groups. If subjects similar to those in this experiment consistently perform in this manner when similar methods of testing are used, then the use of the S.R.A. Self-Scorer method would not be any more justified than the traditional method.

SUMMARY

The purpose of this study was to determine the effect of the S.R.A. Self-Scorer upon learning when used in a testing situation as compared with the use of standard answer sheets when the latter method of testing is accompanied by class discussion of the test items.

Two criteria were used in the measurement of the outcomes of this experiment. Both criteria consisted of objective type examinations administered four weeks after the original learning experience.

The first criterion was the gain achieved on the second administration of a seventy-four-item test and by use of this criterion in the analysis, the results of the investigation indicated that the control group performed significantly better than the experimental group at the one per cent level of confidence. The analysis of covariance used in the analysis yielded an F value of 13.93 with 1 and 207 degrees of freedom.

An analysis of regression was performed to test the significance of the relationship between the first criterion and the control variables. The F value of 1.11 with 3 and 208 degrees of freedom provided evidence that the control variables were not effective in explaining a significant portion of the variation in the gains criterion.

The second criterion was the scores on a twenty-five-item test constructed around concepts measured by the twenty-five most difficult items of the seventy-four-item test. With this post-test as a criterion, analysis of the results showed no significant difference between the experimental and the control groups. Analysis of regression yielded a value of F of 6.90 with 3 and 208 degrees of freedom, which provides evidence that the control variables were effective in eliminating a significant portion of the criterion variation attributable to individual differences used as controls in the experiment.

An analysis of items of the seventy-four-item test was performed comparing responses on both administrations of the test. The results of this analysis indicated that the S.R.A. Self-Scorer does not significantly affect learning on items measuring one type of learning outcome (e.g. definition, knowledge of facts or application of principles) more than on items measuring other types of learning outcomes.

REFERENCES

- 1) George W. Angell, "The Effect of Immediate Knowledge of Quiz Results on Final Examination Scores in Freshman Chemistry," *Journal of Educational Research*, 42: 391-394, 1949.
- 2) George W. Angell and Maurice E. Troyer, "A New Self-scoring Device for Improving Instruction," *School and Society*, 67: 84-85, 1948.
- 3) Donald F. Blood, "The Effect of the S.R.A. Self-Scorer in the Measurement of Spelling Ability," Unpublished doctoral dissertation, State University of Iowa, 1951.
- 4) H. L. Jones and M. O. Sawyer, "A New Evaluation Instrument," *Journal of Educational Research*, 42: 381-385, 1949.

5) James K. Little, "Results of the Use of Machines for Testing and Drill upon Learning in Educational Psychology," *Journal of Experimental Education*, 3: 45-49, 1934.

6) J. C. Peterson, "A New Device for Teaching, Testing and Research in Learning," *Transactions of the Kansas Academy of Science*, 33: 41-47, 1930.

7) Sidney L. Pressey, "A Simple Device Which Gives Tests and Scores—and Teaches," *School and Society*, 23: 575-576, 1926.

8) J. M. Stephens, "The Effect of Symbolic Punishment upon Strong and upon Weak Associations," *The Journal of General Psychology*, 25: 177-185, 1941.

9) Maurice E. Troyer and George W. Angell, *Manual for the S.R.A. Self-Scorer*. Chicago, Science Research Associates, 1949.

EFFECT OF TEACHER-STUDENT INTERVIEWS ON CLASSROOM ACHIEVEMENT¹

ARTHUR J. HOEHN and ELI SALTZ

Air Force Personnel and Training Research Center,

Chanute Air Force Base, Illinois

Many psychologists and educators have encouraged teachers to take on counseling-interview functions (1-3, 6, 11). Few studies are reported in the literature, however, to indicate the effectiveness of such interviews. At present there is very little evidence concerning what kinds of interviews are effective for improving classroom performance, or what types of students benefit most from interviews. In addition, the studies reported in the literature usually employ the experimenter himself as interviewer. It is probably a safe assumption that most experimenters in this area are particularly interested in the effect of personal, teacher-student relations, and are consequently good interviewers.

The present series of studies was designed to investigate the following questions:

(a) Can the average teacher in an Air Force technical training course improve student performance by means of interview techniques?

(b) Can Air Force students be identified who will benefit most from an interview with their teacher?

(c) Can particularly effective interview techniques be identified?

Two of the better controlled experiments reported in the literature are those of Sherriffs (8) and Bradt and Duncan (3). Both studies report success in raising student grades by means of interview techniques. Both of these studies involved only the experimenters as interviewer. Sherriffs found that interviews were most effective with students rated as being high in "self tension," "family tension," and "social tension" on the basis of the interview material.

¹ This research was carried out under the Air Force Personnel and Training Research Center, Chanute Air Force Base, Illinois, in support of Project 7714, Task 77249. The writers wish to express their appreciation to Dr. Henry Adams, Mr. Charles E. Zeleny, and M. Sgt. Robert C. Evans for their assistance in designing and implementing the present research.

Bradt and Duncan utilized a thirty-minute interview, Sherriffs a sixty-minute interview. Bradt and Duncan characterized their most effective interview procedure as one in which the interview was oriented about a discussion of the student's goals and ambitions, in contrast with discussion of world affairs or no interview. Sherriffs characterized his most effective interview procedure as one in which the interview was oriented around the student's life history, interests, and background.

In a somewhat related study, Briggs and Roe (4) report that encouraging airmen to "gripe" had a beneficial effect on the morale of the airmen.

Experiment I attempted to establish a teacher-student relationship similar to that described by Sherriffs, Bradt and Duncan, and Briggs and Roe. As in the Bradt and Duncan and the Sherriffs studies, teachers attempted to discuss the personal problems of the students. As in the Briggs and Roe study, students were encouraged to "gripe" about problems that bothered them. The writers hypothesized that anxiety and rigidity (as defined by tests described later in this paper) should reflect the types of tension described by Sherriffs. Therefore, anxious and rigid students should be helped by the catharsis opportunities of the interviews more than non-anxious and nonrigid students.

EXPERIMENT I

Method

Subjects. Forty-eight students were selected for interviewing out of a group of two hundred and seven students in their first four weeks of the Air Force Weather Observers School, the remaining one hundred and fifty-nine students comprised the control group. The two hundred and seven students comprised two classes, approximately equal in size, which entered the course within four weeks of each other. The course is a complex, technical one, and the majority of students in it are college caliber.

Each entering group was under a different set of two head teachers who did a large part of the teaching. These four men did all the interviewing in Experiment I.

Students chosen for interviewing were randomly selected from a population which had been stratified with regard to anxiety and rigidity. Anxiety was defined by the Taylor Manifest Anxiety Scale (9), a questionnaire device composed of MMPI items selected

by clinicians as indicators of manifest anxiety. Most of the items deal with symptoms of overt anxiety, e.g., sweating. Rigidity was defined by the Wesley Rigidity Scale (10), constructed in a manner similar to that used by Taylor. The items on this scale deal largely with persistence, constriction, and moralistic behavior patterns. Students were considered anxious or rigid if they obtained scores at or above the 80th centile on the respective tests. Students obtaining scores below the 80th centile were considered nonanxious or nonrigid.

Teacher training. The teachers were trained in the interview technique in two groups of two. Each group was trained for one hour on each of three days. On the first day of training, teachers were oriented as to the sorts of interviews they were to conduct. An opening paragraph for use with students was supplied to the teachers in which the teacher was to point out that, due to the size of the classes, it was difficult for teachers to get to know students. Teachers were then to ask questions about the student's marital status, how far from home the student was, and how the student liked the course. When a comment by the student suggested a problem area, this area was to be probed. Following this, the teachers were to follow the students in whatever topics the students raised. Teachers were told to encourage students to gripe about conditions on the base, at home, etc., and to take a permissive attitude toward such griping. During the remainder of the first training session, a psychologist demonstrated the technique using an airman assigned to AFPTRC as interviewee. The demonstration was then discussed with teachers to point up crucial facets of the interview. On the second and third days the teachers practiced interviewing AFPTRC airmen and these practice interviews were discussed.

Interviews lasted approximately thirty minutes and were conducted within the first two weeks of the course for most students. All were completed before the end of the third week.

Criteria. Grades for the first phase of the course were determined on the basis of an objective test given at the end of the fourth week of the course. Because of the high standards set for this course, approximately twenty-five per cent of the students in every entering class fail the first phase of the course and must repeat the phase. The decision on failures, in the two classes studied, was

made by an administrative officer who did not know which students had been interviewed, and was made on the basis of phase grades.

Results

An analysis of variance using phase grades indicated no differences between interviewed and noninterviewed students, and no greater effect of interviews for anxious or rigid as opposed to non-anxious or nonrigid, respectively.

Analysis of failure rates suggests that the interviews may have had some effect on students who were borderline between passing and failing. However, the results are not all in the predicted directions.

It had been anticipated that interviewed airmen would do better in classwork than airmen who were not interviewed. Failure data do not show any *general* tendency toward less failure on the part of interviewed as compared with not interviewed airmen. Interview effects seem to have been specific to personality types. For example, rigid students who were interviewed actually had a *higher* washback rate than those who were not interviewed. This difference, shown in Figure 1, is significant at only about the 0.10 level. As can also be seen in Figure 1, the data do not sustain the prediction that interviews would be more beneficial for rigid than for nonrigid

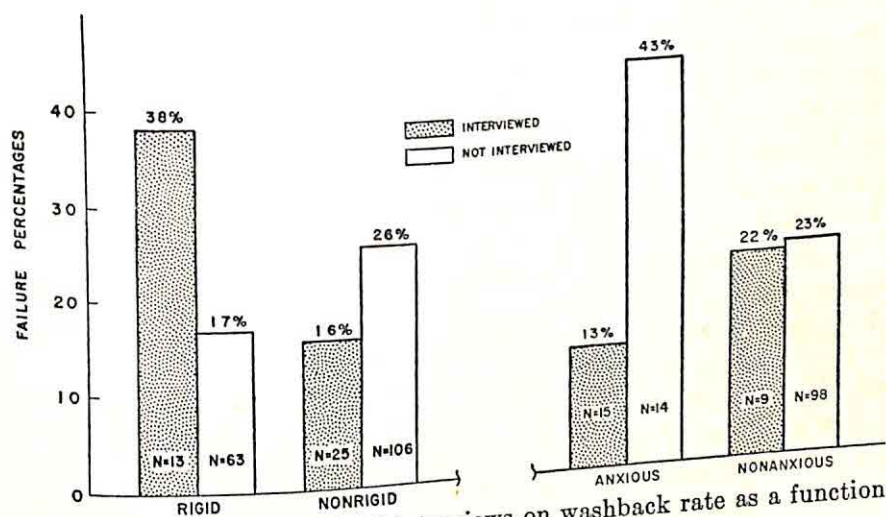


FIGURE 1. Effect of "gripe" interviews on washback rate as a function of rigidity and of anxiety.

students; the difference between the increase in failure rate for interviewed rigid students and the decrease in washback rate for the interviewed nonrigid students is significant at beyond the 0.05 level. These results suggest that rigid students should probably not be interviewed, at least not by the procedures described in this experiment.

The anticipation that interviews would be beneficial to anxious students tends to be sustained, as can be seen in Figure 1. The percentage of failures among interviewed anxious students is considerably less than the percentage of failures among not interviewed anxious students; the difference between percentages was significant at beyond the 0.06 level.

EXPERIMENT II

This experiment was planned as an extension of Experiment I. In the present experiment a larger number of teacher interviewers was employed and a second interview technique was developed with the hope it would prove effective for rigid students. Also, a more complex, factorial design was used in this experiment than in the first; this should increase the sensitivity of statistical analysis.

Clinical theory suggests a possible explanation for the opposite effects of "gripe" interviews on anxious students and rigid students. For anxious students the gripe interviews might have had a catharsis effect that reduced anxiety. However, clinical theory suggests that rigid persons have as a central problem the conflict over their aggressive impulses and that expression of aggression floods them with anxiety. Consequently, "gripe" interviews might actually have increased the anxiety level of rigid students. Therefore, two types of "interview" procedures were used in Experiment II. Some students received the "gripe" interviews of Experiment I, others received a "satisfactions" oriented interview. The satisfactions interviews concentrated on the present sources of satisfaction, the personal goals, and the hopes for the future of students.

Method

Experimental design. Four variables were manipulated in a $3 \times 3 \times 3 \times 2$ factorial design with three students in each of the fifty-four cells. The independent variables were:

- (a) Anxiety: Students were stratified into three levels of anxiety

on the basis of scores on the Taylor Manifest Anxiety Scale (9). High anxious were students at or above the 80th centile. Medium anxious were students between the 21st and 79th centiles. Low anxious were students at or below the 20th centile. It should be noted that the low anxious group in Experiment I was the equivalent of the combined middle and low anxious groups in Experiment II.

(b) Rigidity: Students were stratified on the basis of the Wesley scale (10) into three levels of rigidity similar to the three levels of anxiety. As was true of anxiety, the low rigid group in Experiment I was the equivalent of the combined middle and low rigid groups in Experiment II.

(c) Interview technique: Students were placed in one of three interview groups—"gripe," satisfaction, or not interviewed.

(d) Aptitude: Students were stratified into high and low aptitude for the course on the basis of the Electronic Specialty Aptitude Index. Students with stanines of 6 and 7 were in the low group; those with stanines of 8 and 9 were in the high group.

Subjects. Students were one hundred and sixty-two airmen in the first phase of the Air Force Armament System Fundamental course, a course primarily concerned with fundamentals of electronics. These one hundred and sixty-two students were selected from approximately six hundred entering students. Selection was random within the stratifications determined by the individual difference variables being studied and was accomplished before the beginning of the first phase of the course. All students were above the 70th centile on the aptitude index used, and so the group aptitude is probably not lower than that found in the freshman class of most large state universities.

The interviews were conducted by thirty-one teachers. Teachers interviewed only their own students. Assignment was random with the restrictions that not more than five students requiring interviewing were to be assigned to any one teacher and that no teacher was to have more than one student from any of the fifty-four cells in the design. These restrictions were necessary in order to randomize teacher differences in interviewing and teaching ability. Teachers differed greatly in background. Approximately half were civilian and half were airmen.

Interviews. All interviews took place within the first three days of the first phase.

Each teacher attended these training sessions, conducted as follows.

Session 1: This was a lecture and discussion period approximately one hour in duration and covering the background, purpose and general nature of the study, the rôle which the teachers would play in the study, the kinds of interviews to be given, and the nature of the subsequent training. For this session, teachers met in groups of fifteen to twenty.

Session 2: This session, which teachers attended in groups of four to eight, included a detailed description of the interview procedures, a demonstration, and practice in interviewing. The teachers were told that, in interviewing, they should: (a) create the impression of being genuinely interested in the student, (b) induce the student to choose the topics dealt with and to speak freely in relation to them, (c) exert influence on the content of discussion only insofar as necessary to maintain the differentiation between "gripe" and "personal interest" interviews, (d) adopt an acceptant attitude but avoid expressions of personal approval or disapproval of the students' views, (e) give no advice and make no promises of action aimed at solving the student's personal problems, and (f) conclude the interview on a supportive note.

Following this description of the interviewer rôle, a demonstration was given of both the "gripe" and the "personal interest" interview with teachers taking the rôle of students being interviewed. Subsequent to this, each teacher interviewed another teacher and his performance was critiqued by other members of the group.

Session 3: The third session was a "refresher" session which teachers attended not more than three days prior to the time they did their first interviewing in the experiment. The details of the interviewing were again discussed and any further questions answered. It was again emphasized that in the "personal interest" interview, teachers were to encourage subjects to talk about things they liked, and that in the "gripe" interviews, teachers were to encourage subjects to talk about things that were bothering them.

Two check lists were constructed as devices for obtaining information which teachers were to use in conducting the interviews. On one of these, students were asked to check or list their five principal interests or sources of satisfaction. On the other, students were asked to check or list their five chief "gripes."

All the teacher training was accomplished by one airman research technician. This research technician, an airman with extensive interviewing experience, had been intensively trained by the experimenters in conduct of the specific types of interviews used in the present study.

Interviews lasted thirty to forty-five minutes per student.

Criteria. An objective test of forty-five items was administered to students at the end of the first two-week phase of the course. The questions involved computation of voltages, voltage drops, resistances, etc. The split half reliability of the test was found to be 0.90.

No failure data were collected since only between five and ten per cent of students in the first phase of this course failed it.

Results

As in Experiment I, there is no evidence that interviews significantly affect average grades. The only significant F in the analysis indicated that persons with high aptitude learn more than those with low aptitude. This difference is significant at beyond the 0.005 level, $F = 9.79$ with 1 and 108 degrees of freedom.

While failure rate was too low to permit analysis, it was possible to get some idea concerning the effect of interviews on performance by the borderline achievement students. The lowest score in each cell for low ability students should reflect the performance of borderline students. This analysis reduced the total sample size

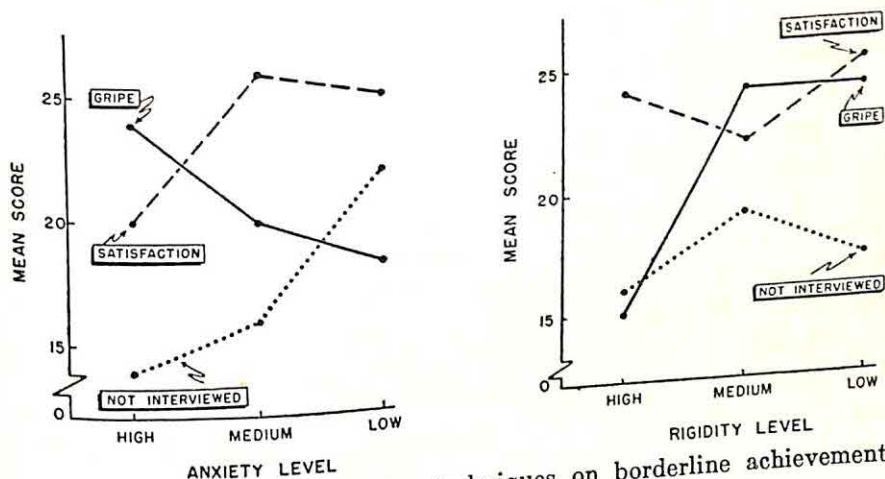


FIGURE 2. Effect of interview techniques on borderline achievement students at various levels of anxiety and of rigidity.

to twenty-seven and statistical analysis was not worthwhile. However, some of the trends may be of interest. Figure 2 indicates the effects of the various interview conditions on the borderline students. Again, as in Experiment I, anxious borderline students tend to be helped by gripe interviews. They are also aided, but to a lesser extent, by the satisfactions interviews. As in Experiment I, there is a tendency for low anxious students to be hindered by "gripe" interviews. There is, however, a tendency for low anxious students to be helped by satisfactions type interviews.

Figure 2 also indicates the effects of interviews on borderline achievement rigid students. As in Experiment I, gripe interviews tend to have a detrimental effect on rigid borderline students. Satisfaction type interviews appear to aid performance of the rigid borderline students.

In general, the two experiments are consistent in not indicating gains, due to interviews, by the average student.

EXPERIMENT III

This experiment was conducted at about the same time as Experiment II. It considers the effect of a group interview technique in which the students were encouraged to gripe. Each group was interviewed twice, as against the one interview in the previous experiment. Personality variables were not considered.

Method

Subjects. Subjects were seventy-seven airmen in phase 1, Mechanical Instrument phase, of the Air Force Aircraft Instrument Mechanics course. Students were randomly assigned to interviewed and not interviewed conditions. The six teacher-interviewers were selected from those regularly assigned to teach the course.

Each teacher interviewed one group of seven of his students. The six interviewed groups had a total of forty-two students. The control groups varied in size from four to seven students, the total being thirty-five. The aptitude level of these students was probably equivalent to a senior high school level.

Interviews. Three one-hour classes in nondirective group counseling techniques were held with the six teachers. Each teacher received a mimeographed description of the proposed study which also outlined his counseling duties. A "model gripe session" was

TABLE I—EFFECT OF GROUP INTERVIEWS ON GRADES

	Interviewed	Not Interviewed
Mean grade	47.30	46.40
S.D.	7.42	8.71
N	42.00	35.00

presented at the first class, followed by a short lecture on the group session procedure.

Both handout and lecture emphasized that the students should be encouraged to express complaints and resentments, especially those related to the course. The teachers were cautioned that they must be permissive and objective and must convince the students that no retaliation was possible.

The last two training sessions consisted of rôle playing. Each teacher was given an opportunity to act as discussion leader while the other teachers took the part of students. The sessions were tape recorded and these recordings were played back to the teachers and analyzed.

Each student group received one interview session during the first week and one during the second week of the three week phase. Each of these lasted approximately forty-five minutes.

Criteria. The measure of achievement in this study was the final phase examination. The Kuder-Richardson (Formula 20) reliability of this test was 0.76.

The failure rate in this course was too small to be of use as a criterion measure.

Results

The difference between phase grades for interviewed and non-interviewed students was small and not statistically significant, as can be seen in Table I.

SUMMARY AND CONCLUSIONS

Previous studies have reported that personal interviews between teachers and their students have resulted in significant average student grade increments. These results have been particularly striking when the interviews dealt with the personal interests and problems of the students. However, in most of the studies reported

the teacher-interviewer was a person whose background made him exceptionally suitable for conducting such interviews. The present series of three experiments tried to reproduce the facilitating results of teacher-student interviews using, as interviewers, a random sample of Air Force teachers. These teachers were given a three hour course in interview training that included lectures and rôle playing. The interview techniques taught were based on those that previous studies have reported as particularly effective.

The results suggest that the interviews influenced failure rates. However, this influence was not always beneficial. The results suggest that while "*gripe*" oriented interview techniques may *decrease* the failure rate for anxious students, these techniques may actually *increase* the failure rate for rigid students. On the other hand, interviews oriented toward *personal sources of satisfactions* appear to *decrease* the failure rate among borderline achievement rigid students. These results suggest that teacher interviews may be harmful in some cases.

All three experiments were consistent in finding that the *average grades* of interviewed students were not better than the average grades of students who were not interviewed. These results were consistent for *individual* interviews oriented toward student problems and "*gripes*;" for *group* interviews oriented toward student gripes; and for *individual* interviews oriented toward student goals, hopes for the future and sources of personal satisfaction. Nor did the student personality variables examined interact with interviews by the random sample of teachers. Interviews did not raise grade averages for anxious, rigid, nonanxious, or nonrigid students. It should be noted that Martin (?), too, found that interviews did not affect average grades but did affect number of course failures.

These results suggest that the average teacher, with only a minimal amount of training in interview techniques, cannot improve grade averages by means of interviews. Coupled with the possibility that such interviews may be harmful for some borderline achievers, these results indicate that caution should be exercised in recommending that teachers counsel their students.

Definite implications for future research emerge from these experiments: (a) The criterion of failure rate should be considered more carefully than it has in the past. More emphasis should be placed on the effects of interviews on borderline achievers. (b)

The sorts of teachers who can succeed in interviewing should be determined. Just as different interview techniques have different effects on students, depending on students' personality, teacher personality may influence the type of interview technique which a teacher can use successfully. It would be consistent with clinical intuition if future results showed that specific teachers were successful interviewers only when they used specific techniques, and used them only on students with specific personality characteristics. (c) The interaction between student personality and specific interview techniques should be investigated.

REFERENCES

- 1) D. S. Arbuckle, "Can English Teachers be Counselors?" *English Journal*, 42: 192-193, 1953.
- 2) D. S. Arbuckle, *Teacher Counseling*. Cambridge, Mass., Addison-Wesley Press, 1950.
- 3) K. W. Bradt and C. P. Duncan, "Degree of Personal Relationship between Instructor and Student as a Factor in Course Grade Improvement," *American Psychologist*, 6: 368 (abst), 1951.
- 4) L. J. Briggs and R. M. Roe, "Morale as a Function of Opportunity to Register Complaints," *Technical Report HRRC-TR-53-4*, Human Resources Research Center, Lackland Air Force Base, Texas, March, 1953.
- 5) C. A. Curran, "Guidance and Counseling in Education," *Education*, 73: 223-228, 1952.
- 6) C. S. Knappe and E. A. Walsh, *Training Analysis and Development Information Bulletin*, 3: 8-12, 1952.
- 7) G. C. Martin, "Interviewing the Failing Student" *Journal of Educational Research*, 46: 53-60, 1952.
- 8) A. C. Sherriffs, "Modification of Academic Performance through Personal Interview," *Journal of Applied Psychology*, 33: 339-346, 1949.
- 9) Janet A. Taylor, "A Personality Scale of Manifest Anxiety," *Journal of Abnormal and Social Psychology*, 48: 285-290, 1953.
- 10) Elizabeth Wesley, "Perseverative Behavior in a Concept-Formation Task as a Function of Manifest Anxiety and Rigidity," *Journal of Abnormal and Social Psychology*, 48: 129-134, 1953.
- 11) G. C. Wrenn, *Training Analysis and Development Information Bulletin*, 4: 7-12, 1953.

A LONGITUDINAL STUDY OF SIBLING RESEMBLANCES IN INTELLIGENCE AND ACHIEVEMENT¹

SARAH M. SCHOONOVER

The major problem of this study is to determine the amount of sibling resemblance in longitudinal growth records of mental ability and educational achievement. The following specific points have been investigated: (a) To what extent do children of the same family resemble each other in intelligence and in educational achievement? (b) To what extent are differences between siblings reduced as compared to those that exist between unrelated pairs within the same population? (c) To what degree do the sibling resemblances in intelligence and achievement correspond to each other?

Investigations concerning familial resemblances in mental ability have been numerous since the initial studies were made by Galton nearly ninety years ago. Studies of the resemblances of siblings in mental ability and in school achievement have been reviewed in the writer's doctoral dissertation (2). The present study has utilized the longitudinal approach to this problem in a manner that has not been employed with data analyzing sibling resemblances in intelligence or achievement.

SIBLINGS SELECTED FOR STUDY

Source of the data. The data were secured from the records of the University Elementary School at the University of Michigan. Of the children in this school, seventy per cent come from professional homes, while thirty per cent come from business and white-collar homes.

There are approximately seven applications for each vacancy at the University Elementary School. The primary purpose for which the school was established, i.e., the promotion of research

¹ This report is a portion of the writer's doctoral dissertation, "Sibling Resemblances in Achievement" (Supplemented by the mental-age data). See: *Dissertation Abstracts*, Vol. 13, No. 5, Publication No. 5726, Ann Arbor, University Microfilms, 1953. Acknowledgment is made to Dean Willard C. Olson, University of Michigan, who directed the original research. The author's present location is Ottawa, Ohio.

in child development, is the main guide in admission policies. Consideration is given to the suitability of children as subjects for growth studies and for the type of educational program provided. Applications are not accepted for the blind, the deaf, the seriously crippled, nor the severely mentally retarded. Preference is given to children likely to remain through sixth grade, to children seeking entrance at an early age, and to children having siblings already enrolled. An attempt is made to keep a balance between the number of boys and the number of girls at each age level.

The school endeavors to provide for the development of the child as a whole in his social setting. It takes into account physical hygiene, mental hygiene, pacing of growth, individual differences, and active participation in group life. Its curriculum is flexible and emergent, is based upon a concept of learning through experience, and is determined in the light of children's interests and present needs.

Description of the data. The psychology division of the University Elementary School maintains an intelligence test schedule on a yearly basis. The Stanford-Binet Revised Scale is administered annually within two weeks of each child's birthdate, or six months following the birthdate of children born in the summer. An achievement test schedule is maintained on a six-month basis. The Gates Reading Test is given semiannually in October and in April, until the child achieves a score of 102 months, at which time he is given the Stanford Achievement Test, Primary Battery. The Intermediate Battery is administered in grades four through six.

Records of intelligence and achievement test scores have been filed in the Psychometric Unit for all children who have attended and who have been tested at the University Elementary School. Since the school was opened in 1930, longitudinal material has been collected over a twenty-five-year period.

Selection of the data. All true sibling pairs, with chronological age overlap, and with four or more scores per sib on the Stanford-Binet Test and on the Stanford Achievement Test, from the fall of 1929 through the spring of 1951, were utilized in this study. With these qualifications, fifty-nine sibling pairs were found for intelligence; sixty-four pairs for arithmetic, education, reading, and spelling; forty-two pairs for literature and social studies; forty pairs for language; and thirty-eight pairs for science. There were fewer sibling combinations in language, literature, science, and social

studies because these tests are not found in the Primary Battery; therefore, there is less chance for a child to have four scores on these tests than for him to have that number on the other tests in the battery.

METHODS FOR THE COMPARISON OF LONGITUDINAL SIBLING RECORDS

The longitudinal approach. The longitudinal records, which involved measurement of the same children year after year, provided the data for this study. They have the following advantages: (a) Some of the variables found in the test situation, such as the child's disposition on any particular date or the child's reaction to being tested for the first time, are minimized. (b) Comparison of children at the same chronological age, rather than at dissimilar ages, is permitted.

For each family included in this study a mental growth graph and eight achievement growth graphs were constructed. This meant that three hundred and forty-four growth graphs and seven hundred and fifty-seven individual growth curves were plotted.

Chronological, mental, and achievement ages were expressed in months. Chronological ages were plotted on the abscissas of the graphs, while mental and/or achievement ages were plotted on the ordinates of the graphs.

Fitting the linear equation. The linear equation best fitting the data was found to eliminate the observed variation and to determine some constant rate of growth which may be used to characterize the observed results. The equation of a straight line, $y = ax + b$, which gives the slope and the intercept of the line used to describe the growth-age relationship, was found by the method of the least squares fit. The linear fit for each child for intelligence and for each of the eight achievement variables was plotted graphically.

The method of the mean of the average differences. For each pair of siblings the limits of the overlap of their chronological ages were found, and from these the midpoint of the overlap was computed. Using the linear best fit, the age scores of each pair of siblings were read at these midpoints. The difference between these two ages was found, yielding the average difference for each sibling pair. The sum of these average differences was divided by the number of pairs, to secure the mean of the sib average differences. In addition to the group of siblings as a whole, the above was calcu-

lated for brother-brother, sister-sister, and brother-sister combinations.

It may be noted that the difference found for each pair is the average difference for the whole overlap period for the pair, because of the nature of the straight line. Also, because the linear best fit was found by utilizing a number of test scores for each child, a longitudinal effect is reflected in the average differences.

The unrelated pairs. To secure comparable data for unrelated pairs, siblings of this study were paired randomly with a non-related partner, and the means of the average differences were calculated for the unrelated pairs. By utilizing the same children for the sibling and for the nonsibling pairs, all group variables (such as age, sex, IQ, EQ, physical factors, emotional factors, social maturity, home environment, and school environment) are neutralized, resulting in two groups, comparable to each other in all of these respects.

The ratios. Ratios were computed by using the means of the average differences of the sib-pairs as dividends and those of the corresponding unrelated pairs as divisors. The resulting ratios were subtracted from 1.00 and then were multiplied by one hundred, producing in percentages the extent to which the sibling differences were reduced beyond those of the nonrelated pairs. To determine the significance of the differences between the sib-pair means and the unrelated pair means a *t* test for significance was employed.

The method of correlation. In addition to the ratio of the means, correlation coefficients were computed for the sibling pairs. A Pearson product-moment formula was used, as was the method of intraclass correlation. To discover the significance of these correlations a *t* test of an observed correlation was employed.

SIBLING DIFFERENCE AND RESEMBLANCE

The mean of the average differences. The results obtained by the method of the mean of the average differences for sibling pairs and for unrelated pairs are given in Table I for intelligence and achievement.

The percentages of the reduction of the means of the average differences between the sibling and the nonrelated pairs in this study are given in Table II.

These siblings consistently were found to have smaller means of average differences than the unrelated children for the group

TABLE I—MEANS OF AVERAGE DIFFERENCES IN MENTAL AND EDUCATIONAL AGE SCORES, EXPRESSED IN MONTHS

Mental or Achievement Age Measure	Total Group	Boy-Boy	Girl-Girl	Boy-Girl
Mental				
Sibling	9.9	9.4	8.1	11.5
Unrelated	18.3	23.5	14.2	18.1
Arithmetic				
Sibling	10.7	12.2	9.2	10.6
Unrelated	15.1	17.7	10.1	13.8
Education				
Sibling	12.6	12.2	12.3	13.0
Unrelated	19.3	23.6	17.7	15.8
Language				
Sibling	20.7	16.1	23.3	22.2
Unrelated	30.1	31.8	36.5	28.5
Literature				
Sibling	14.3	10.0	16.7	15.6
Unrelated	20.7	18.2	22.7	17.4
Reading				
Sibling	16.4	13.0	17.8	18.3
Unrelated	25.9	32.6	22.4	26.1
Science				
Sibling	16.0	14.3	16.1	16.7
Unrelated	20.5	24.1	23.3	23.0
Social studies				
Sibling	13.6	12.3	9.7	17.6
Unrelated	19.7	25.2	13.8	21.8
Spelling				
Sibling	14.7	13.1	14.3	16.1
Unrelated	21.5	25.5	20.3	20.8

TABLE II—PERCENTAGES OF REDUCTION OF MEANS OF AVERAGE DIFFERENCES BETWEEN SIBLING AND NONSIBLING PAIRS

Mental or Achievement Age	Total Group	Boy-Boy	Girl-Girl	Boy-Girl
Mental	45.6	60.4	42.8	36.5
Arithmetic	29.2	31.1	8.9	23.2
Education	34.8	48.3	30.5	17.8
Language	31.2	49.4	36.2	22.1
Literature	31.0	45.1	26.5	10.4
Reading	36.7	60.1	20.6	30.0
Science	22.0	40.7	31.0	27.4
Social studies	31.0	51.2	29.8	19.3
Spelling	35.4	52.6	42.0	25.6

as a whole, as well as for boy-boy, girl-girl, and boy-girl combinations. The greatest amount of reduction in variation, in each measure, was found in the boy-boy pairs. The question arises: are brother-brother combinations more alike than the other groupings, or are the differences between unrelated boy-boy pairs merely greater? The means of the average differences indicated that the differences between unrelated boys were greater than those of the other groupings, and that the brother-brother pairs showed only slightly less variation than what was found for the total group.

The smallest amount of reduction in variation, 8.9 per cent, was for girl-girl combinations in arithmetic. Girls as a whole in this study were more similar to each other in arithmetic ability than they were in other achievement-age measures. The largest amount of reduction in variation, 60.4 per cent, was found for boy-boy combinations in mental age. Unrelated boys varied to a greater extent in mental ability than did the other groupings, and brothers were slightly less variable than was the total group.

In general, the content subjects showed less reduction in variation than did the skill subjects. Perhaps this was caused partly by the greater unreliability of tests in the content subjects. All the differences between the sib-pair means and unrelated pair means for the total group were significant at the five per cent level or lower, except for science. The number of sibling and unrelated pairs in science was the smallest of all the achievement measures, which may explain in part why the differences of the science means, though in the expected direction, did not meet the same test for significance.

TABLE III—CORRELATIONS FOR THE TOTAL GROUP
OF SIBLING AND NONSIBLING PAIRS

Mental or Achievement Age	Siblings	Nonsiblings
	0.71	0.27
Mental	0.49	0.08
Arithmetic	0.59	0.02
Education	0.40	-0.09
Language	0.41	-0.11
Literature	0.51	0.01
Reading	0.39	-0.02
Science	0.64	0.14
Social studies	0.53	0.08
Spelling		

The correlation coefficients. The results secured by the method of correlation coefficients for the total group of sibling pairs and for the total group of unrelated pairs are given in Table III. For intelligence the correlation coefficient found for the siblings was high, indicating a marked relationship. The correlation for the nonrelated pairs was low, indicating a small relationship. Theoretically, the correlation expected for the unrelated pairs is zero. The fact that approximately ninety-five per cent of the children in this study have mental growth curves above the so-called norm may, in part, be responsible for the correlation for the unrelated pairs yielding a small relationship, rather than no relationship.

For achievement the correlation coefficients for the siblings were moderate, indicating a substantial relationship. The correlations found for the nonrelated pairs average $+0.01$ for the eight achievement measures.

According to the obtained t values, all the sibling correlations were significantly different from zero. It may be pointed out that these correlations substantiate each other and give some indication of the value of the correlations in the population, since they are all positive and are in a relatively narrow range.

The children involved in this study were a select and homogeneous group, since the large majority of them came from homes of superior economic, social, and intellectual status. It is possible that the differences between the unrelated pairs were smaller than they would have been if the total group had been less homogeneous.

CONCLUSIONS

All three methods of analysis utilized on the longitudinal growth records in this study, i.e., means of the average differences, percentage reduction of difference by family membership, and correlation, produced consistent results in describing the existence of a substantial amount of sibling resemblance in intelligence and achievement. Resemblances in intelligence were somewhat greater than they were in achievement.

REFERENCES

- 1) Willard C. Olson and Byron O. Hughes, *Manual for the Description of Growth in Age Units*. Ann Arbor, Edwards Letter Shop, 1950.
- 2) Sarah M. Schoonover, *Sibling Resemblances in Achievement*. Ann Arbor, University Microfilms, 1953.

BOOK REVIEWS

ARNOLD BERNSTEIN. *On the Nature of Psychotherapy*. New York: Doubleday & Company, pp. 36, 1954.

EMERY I. GONDOR. *Art and Play Therapy*. New York: Doubleday & Company, 1954, pp. 36.

ROBERT I. WATSON. *Psychology as a Profession*. New York: Doubleday & Company, 1954, pp. 65.

These are part of a series of Doubleday Papers in Psychology. They are paper-bound booklets which attempt to give concise statements about psychology in its various areas. They are written by individuals chosen for special interest and acquaintance with each special topic. They can be read by those who have not had a beginning course in psychology as well as by students of psychology who desire a special discussion of a particular subject.

Bernstein's booklet discusses, among other things, the nature of disease and of therapy, therapeutic acts, instruments and relationships, somatogenesis, psychogenesis, sociogenesis, the therapy group and the psychotherapeutic relationship, the patient's point of view, the point of view of the psychotherapist, the psychology of the therapist, and for further reading gives a well-selected list of books and papers.

Gondor's *Art and Play Therapy* deals in Section 1, with theory and practice; the child in therapy; and art and play in therapy. Section 2 deals with clinical applications: special technique for first contact; preliminary art and play therapy; and art and play therapy with retarded children. Section 3 presents three case histories illustrating first contacts with child and parents, achievement of a limited goal with a deeply disturbed child. There are two pages of selected references.

Watson's *Psychology as a Profession*, is of general interest and value for all who desire to know the problems of psychologists and their work in professional activities. Section 1 gives an orientation, relating to the nature of a profession, the professional concept in psychology and the kinds of specialization in psychology. Section 2 deals with the stabilizing professional factors in psychology: the American Psychological Association, training in professional psychology, ethics in the profession, and various organizations, national and local, that make for professional stabilization. Section 3 takes up unresolved intraprofessional issues: professional vs.

scientific influences, and practice on the subdoctoral level. Section 4 discusses various unresolved interprofessional issues, including the psychologist and psychotherapy, counseling, private practice, legal problems, and the relation of psychology to medicine. An excellent list of one hundred and eighteen references is appended.

These and others of the Doubleday Papers in Psychology will probably find extensive use for those who desire fairly nontechnical and generally good discussions of such subjects as *The Natural Man*, *Nature and Nurture*, *Feelings and Emotions*, *Learning*, *Individual Development*, etc.

A. S. EDWARDS

The University of Georgia

NELSON B. HENRY, Editor. *Mental Health in Modern Education, Part 2, Yearbook of the National Society for the Study of Education*. Chicago: The University of Chicago Press, pp. 397, lxxiv.

Under the chairmanship of Paul A. Witty and a committee consisting of H. A. Carroll, P. T. Rankin, H. N. Rivlin and Ruth Strang, and twelve associated contributors, a valuable and for the most part well-condensed study is here presented on the problems of mental health in modern education.

Although rather long delayed as a major subject of examination in this series of studies, it is very timely. It is one of several recent publications on the subject which many believe to be more than half of the whole problem of health.

The importance of mental health is emphasized in a number of places in the volume. One such emphasis is expressed by Shibler (pp. 303-4) who writes: "The Citizenship Education Study in Detroit points to one conclusive fact: The basis for training for citizenship in the schools is good mental health."

This comprehensive study deals with many phases of the subject under the following large heads: the needs for the program and the rôle of this yearbook in the improvement of mental health; the history and present status of the mental health movement; conditions affecting mental health in the classroom; problems and practices related to mental health at different educational levels; personal and professional development of the teacher; and mental health for today and tomorrow.

In more detailed analysis we find such subjects as motivation and learning in mental health; the rôle of the home, of the school, and of the community in the mental health program; special problems at the preschool level, in the primary grades, in the intermediate grades, in the high school, at the college level, and an evaluation of present programs.

Some brief discussion is given of the contributions of anthropology, sociology, psychology, general semantics, philosophy, religion and law. Reference is made to mental mechanisms, psychoneuroses, emotional security, emotional conflict, counseling, causes of mental and emotional deviation, adjustments, neurotic and psychotic types.

Many but not all of the chapters have excellent lists of references at the ends; a large number of footnotes accompany some of the chapters.

The great comprehensiveness of the study makes impossible the more detailed discussions which many will want, but which they can find in the special articles and books on the various subjects. There is an index of six pages which is well made but not quite as complete as it might well be.

The appendices (pp. i-lxxiv) present the constitution and by-laws; minutes of the Atlantic city meeting; synopses of some proceedings of the Board of Directors; report of the treasurer; a list of members; information concerning the society; and a list of the publications of the society.

The book is a very valuable contribution to a most important subject—one of the most important for parents and teachers. It gives evidence that the lives, efforts and hopes of such men as Clifford Beers and W. H. Burnham were not in vain.

A. S. EDWARDS

The University of Georgia

PUBLICATIONS RECEIVED

- VERGILIUS FERM AND OTHERS. *A Dictionary of Pastoral Psychology*. New York: Philosophical Library, 1955, pp. 336. \$6.00.
- Fifty-Fifth Annual Report of the Superintendent of Schools City of New York: Statistical Section, School Year 1952-1953*. Prepared by the Bureau of Administrative and Budgetary Research. Board of Education, New York City.
- Fifty-Sixth Annual Report of the Superintendent of Schools of New York City*. Prepared by the Bureau of Administrative and Budgetary Research. Board of Education. New York City.
- ARTHUR E. FINK, EVERETT E. WILSON AND MERRILL B. CONOVER. *The Field of Social Work*. Third Edition. New York: Henry Holt & Co., 1955, pp. 630. \$5.25.
- ARTHUR W. FOSHAY, KENNETH D. WANN, AND ASSOCIATES. *Children's Social Values: An Action Research Study*. New York: Bureau of Publications, Teachers College, Columbia University, 1954, pp. 323. \$3.50.
- LAWRENCE K. FRANK. *Individual Development*. New York: Doubleday & Co., Inc., 1955, pp. 52 (paper). \$0.85.
- D. C. FRASER. *A Psychological Glossary*. Cambridge: W. Heffer & Sons Ltd., 1954, pp. 40 (paper).
- FRANK S. FREEMAN. *Theory and Practice of Psychological Testing*. Revised Edition. New York: Henry Holt & Co., 1955, pp. 609. \$5.25.
- ELIOT FREIDSON, EDITOR. *Student Government, Student Leaders and the American College*. Philadelphia, Pa.: United States National Student Association, 1955, pp. 88 (paper).
- KARL C. GARRISON AND J. STANLEY GRAY. *Educational Psychology*. New York: Appleton-Century-Crofts, Inc., 1955, pp. 505. \$5.00.
- FRANK H. GERI. *Illustrated Games and Rhythms for Children. Primary Grades*. New York: Prentice-Hall, Inc., 1955, pp. 196. \$2.95.
- EDWIN E. GHISELLI AND CLARENCE W. BROWN. *Personnel and Industrial Psychology*. Second Edition. New York: McGraw-Hill Book Co., 1955, pp. 491. \$6.00.
- H. HARRY GILES. *Human Dynamics and Human Relations Education*. New York: New York University Press, 1954, pp. 108 (paper).
- JAMES M. GILLESPIE AND GORDON W. ALLPORT. *Youth's Outlook*

- on the Future*. Garden City, New York: Doubleday & Company, Inc., 1955, pp. 61 (paper). \$0.85.
- EMERY I. GONDOR. *Art and Play Therapy*. Garden City, New York: Doubleday & Company, Inc., 1954, pp. 61 (paper). \$0.95.
- Graphic Arts Shop, Teacher's Work Manual*. Board of Education of the City of New York. 1954, pp. 77 (paper).
- HARRY A. GREENE AND HELEN I. STAPP. *Greene-Stapp Language Abilities Test*. Yonkers-on-Hudson, N. Y.: World Book Company (paper).
- JULIEN GREENLEE. *Better Teaching Through Elementary Science*. Dubuque, Iowa: Wm. C. Brown Company, 1954, pp. 204 (paper). \$2.50.
- JULIAN GREENLEE. *Teaching Science to Children*. Dubuque, Iowa: Wm. C. Brown Company, 1955, pp. 185 (paper).
- WILLIAM W. GRINGS. *Laboratory Instrumentation in Psychology*. Palo Alto, California: The National Press, 1954, pp. 282. \$4.75.
- J. E. GRINNELL, RAYMOND J. YOUNG AND OTHERS. *The School and the Community: Educational and Public Relations*. New York: The Ronald Press Co., 1955, pp. 444. \$5.50.
- SIDONIE MATSNER GRUENBERG, Editor. *The Encyclopedia of Child Care and Guidance*. New York: Doubleday & Co., Inc., 1954, pp. 1016. \$7.50.
- A Guide to Some Educational Resources in the City of New York*. Curriculum Bulletin, 1953-1954 Series, Number 7. Board of Education of New York City, 1954, pp. 107 (paper).
- J. P. GUILFORD. *Psychometric Methods*. Second Edition. New York: McGraw-Hill Book Co., 1954, pp. 507. \$8.50.
- A. PAUL HARE, EDGAR F. BORGATTA AND ROBERT F. BALES. *Small Groups*. New York: Alfred A. Knopf, 1955, pp. 666.
- THISTLE Y. HARRIS. *The Teaching of Nature Study*. Melbourne: Australian Council for Educational Research, 1954, pp. 100 (paper).
- Health Supervision of Young Children. A Guide for Practicing Physicians and Child Health Conference Personnel*. New York: American Public Health Association, 1955, pp. 179. (paper) \$2.00.
- NELSON B. HENRY, Editor. *Mental Health in Modern Education. The Fifty-fourth Yearbook of the National Society for the Study of Education*. Part II. Chicago: The University of Chicago Press, 1955, pp. 397 (paper).
- VIRGIL E. HERRICK AND LELAND B. JACOBS. *Children and the Lan-*

- guage Arts. Englewood Cliffs, N. J.: Prentice-Hall Inc., 1955, pp. 524. \$5.50.
- Home Economics, Clothing and Household Fabrics, Grades 7, 8 and 9.* Board of Education of New York City, 1954, pp. 120 (paper).
- THOMAS W. HOPKINS, HARRY V. BICE AND KATHRYN C. COLTON. *Evaluation and Education of the Cerebrally Palsied Child.* Washington, D. C.: International Council for Exceptional Children, 1954 pp. 114 (paper).
- JAMES L. HYMES, JR. *Behavior and Misbehavior.* New York: Prentice-Hall, Inc., 1955, pp. 140.
- JAMES L. HYMES, JR. *A Child Development Point of View.* New York: Prentice-Hall, Inc., 1955, pp. 145.
- ARTHUR T. JERSILD. *When Teachers Face Themselves.* Columbia University: Teachers College, Bureau of Publications, 1955, pp. 169. \$3.65.
- WENDELL JOHNSON AND RALPH R. LEUTENEGGER. *Stuttering in Children and Adults.* Minneapolis: University of Minnesota Press, 1955, pp. 472. \$5.00.
- HARRY W. KARN AND JOSEPH WEITZ. *An Introduction to Psychology.* New York: John Wiley & Sons, Inc., 1955, pp. 315. \$3.90.
- S. DONALD KARL AND BARBARA L. DIEHL. Editors. *The College Handbook.* New York: College Entrance Examination Board, 1955, pp. 399 (paper).
- FRED S. KELLER. *Learning Reinforcement Theory.* Garden City, New York: Doubleday and Company, Inc., 1954, pp. 37 (paper). \$0.85.
- VICTOR H. KELLEY AND HARRY A. GREENE. *Kelley-Greene Reading Comprehension Test.* Yonkers-on-Hudson, N. Y.: World Book Company (paper).
- ED KILMAN AND THEON WRIGHT. *Hugh Roy Cullen: A Story of American Opportunity.* New York: Prentice-Hall, Inc., 1954, pp. 376. \$4.00.
- LEONARD KORNBERG. *A Class for Disturbed Children: A Case Study and Its Meaning for Education.* New York: Bureau of Publications, Teachers College, Columbia University, 1955, pp. 157. \$3.75.
- RUTH KOTINSKY AND HELEN L. WITMER, Editors. *Community Programs for Mental Health.* Cambridge, Mass., 1955, pp. 362. \$5.00.
- W. D. LAMONT. *The Value Judgement.* New York: Philosophical Library, 1955, pp. 335. \$6.00.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Volume 47

December, 1956

Number 8

AN EVALUATION OF A STUDENT-CENTERED AND INSTRUCTOR-CENTERED METHOD OF CONDUCTING A GRADUATE COURSE IN EDUCATION

GLEN R. RASMUSSEN

School of Education, University of Michigan

This study was designed to investigate the effectiveness of two learning situations in achieving the following goals: (1) to help students learn as much as possible about present knowledge in the area of child psychology, and (2) to help students use this knowledge to guide their own behavior. In this situation the mere acquisition of data and generalizations seemed to be the less important of the two goals to the instructor. It is, of course, much easier to evaluate the success of a course from the standpoint of the first goal, acquisition, than from the standpoint of the second goal, the ability and willingness to apply classroom learning to one's everyday performance in dealing with children.

In the light of the instructor's goals for the course, what evidence can be found concerning the relative effectiveness of different teaching methods? Research attempting to measure "acquisition of subject matter" has shown conflicting results. Most of this research has been conducted in general psychology classes. Two major obstacles have been apparent. First, students in the student-centered classes were *told* they were to be student-centered. They had no choice in the situation. This is a rather autocratic and inauspicious beginning for a democratic experience. Second, how do we measure the kind of learning induced by such dissimilar methods with a paper and pencil test? A test designed by the instructor of a student-centered class violates the philosophy upon which the entire class is based. The instructor is once more the "keeper of the knowledge." Some experimenters (1) have felt

this conflict and informed the student-centered groups that this "final exam" would not count toward a grade. If this examination has grade meaning for one group and does not have grade meaning for the other, the factor of differential motivation is introduced and the results are subject to doubt.

The second goal of the instructor (to help students use course knowledge to guide their own behavior) is even more difficult to measure. Research in this area is practically nonexistent. There is considerable evidence that student-centered psychology classes result in significantly greater emotional and social adjustmental gains for students than do instructor-centered classes (1-3, 5-7). In this study no attempt was made to measure adjustmental gains. Student comments both during and after class suggested that the student-centered class was definitely superior in facilitating social adjustment. It would seem logical to assume that if greater emotional and social adjustmental gains were made by members of the student-centered sections in this study, such gains would be accompanied by an increased ability to use modern psychological principles in their own teaching.

METHOD

The data for this study were gathered in thirteen sections of a class entitled "The Psychology of Child Development." These classes were conducted in eleven different communities under the auspices of the Extension Service of a large midwestern university. Each class met once a week for sixteen weeks. Ninety per cent of the students in these classes were school teachers. Eighty-one per cent were graduate students and seventy-four per cent were parents.

Data were obtained through questionnaires filled out at the end of the course and a test and questionnaire administered six months after the course was completed. Neither the questionnaires nor the tests were signed by the students and the grades were established before the instructor saw any of the questionnaires.

In the instructor-centered sections class time was distributed as follows: informal lectures, nineteen hours; films, five hours; small discussion groups, five hours; examinations, three hours. Students were expected to read the text and pass the examinations. The instructor attempted to clarify and expand the material in the text. This was done through informal lectures. Students were

encouraged to ask questions, give opinions, etc. at any time during the class. The instructor halted questions when they strayed outside of the unit covered during that period. The instructor attempted to be informal, friendly, and helpful in meeting student needs. At various times people in the community were brought in as resource people. This was often done at the suggestion of someone in the class. The procedure for evaluation was handled through two comprehensive examinations, both of which consisted of approximately sixty per cent objective type questions and forty per cent essay.

In the student-centered sections class time was distributed as follows: films, three and one-half hours; class discussion, twenty-eight and one-half hours. The students' responsibilities were to read the text and write between fifteen and twenty personal reactions to either items within the text or within the class discussions. This requirement of writing personal reactions was patterned after Nathaniel Kantor's method (4). Students were encouraged to put down their real feelings about class content or method. Actually the only assignments which were returned to the students as unsuitable were those which consisted of either irrelevant material or else an outline of subject matter rather than a personal reaction. Whenever possible in the student-centered classes the entire class sat around a group of tables with the instructor moving from one spot to another from week to week. When a procedural matter was under discussion ("When shall we see films?; How can we best use them?; Should we have a secretary to keep notes of class discussions?" etc.) the instructor attempted to act as discussion leader if no one else assumed that responsibility. During class discussions of the course content, the instructor seldom attempted to stimulate discussion, practically never made evaluatory comments concerning contributions, and during the first half of the course made a strong attempt to refrain from personal opinions on the items under discussion. At no time did the instructor call on a student. If asked for a personal opinion he tried to give a straightforward and honest answer. The instructor felt that the student-centered situation could most effectively promote significant learning by reducing to a minimum the individual threat to the learner, thus encouraging a more differentiated view of the subject matter. A rather common example of this within these classes was the feeling by some individ-

ual teachers that "some kids don't respect you unless you show that you can get tough with them." The holder of such an opinion might believe that others in the class, including the instructor, would look down upon him for holding such an opinion. Such feelings threaten his status with the group. This threat, and perhaps the guilt feelings attached to holding such an opinion, could very effectively reduce the ability of this person to accept other points of view. If he feels threatened by the holding of such an opinion, he must spend his time deciding how to defend himself rather than checking the validity of the opinion itself. The instructor attempted to reduce this individual threat by: (1) accepting all contributions as worthy and sincere; (2) paying full attention to each speaker, not waiting impatiently for him to finish so that the instructor could make his own comment; (3) pointing out that the teacher's rôle is one of the most difficult to play and that because of a difference in our background, etc., each teacher's approach to the classroom may be valuably different. In addition the instructor attempted to link the various comments together—that is, he attempted to point out the connections between individual speaker's contributions. At the end of each session the instructor made no attempt to provide closure. No final answers were given.

Classes smaller than twenty-five were given their choice as to learning method. The instructor presented an outline of the possibilities and requirements of the instructor-centered and student-centered classes and asked the students to discuss this in either small groups or as a whole and reach a decision while the instructor left the room. The four instructor-centered classes ranged in size from twenty-four to fifty-three; the mean class size was thirty-six. The nine student-centered classes ranged in size from eleven to twenty-four; the mean class size was eighteen. The instructor-centered classes were generally held in large industrial cities. These areas pay higher teacher salaries and may get better educated and more intelligent teachers than the smaller, more rural communities. There was a slightly higher percentage of graduate students in the instructor-centered classes. An Otis Mental Abilities Test was administered to two student-centered classes and one instructor-centered class. The latter scored slightly higher than both of the student-centered classes. In both learning situations the instructor attempted to be kind and understanding.

He made no attempt to vary either his own personality or the social climate. Verbal freedom was encouraged in both situations.

Six months after the course was completed a questionnaire and test were mailed to each student. Fifty-three students who had taken part in an instructor-centered course responded. In the six-month interval since the end of the course these fifty-three students had taken a total of thirty-eight courses in education or psychology. Seventy-four students who had taken part in a student-centered class responded. In the same interval they had taken a total of twenty-one classes in education or psychology. $\chi^2 = 25$, sig. < 0.01.

RESULTS AND DISCUSSION

On the basis of formal study in the six-month interval between the completion of the class and the examination, members of the instructor-centered sections should have demonstrated superior scores on the examination. This examination consisted of items previously used in either the mid-semester or final examination in the instructor-centered classes. Therefore the members of these sections were familiar with the items, type of test, etc. Members of the student-centered sections had not taken any tests and therefore found themselves faced with a new situation.

On the basis of four counts—(1) more intervening courses in education and psychology, (2) slightly higher proportion of graduate students, (3) slightly higher mean I.Q., (4) familiarity with test style and test items—it might be expected that members of the instructor-centered classes would achieve higher scores on the test. Actually no significant difference was found. Students who had participated in the instructor-centered classes achieved a mean raw score of 28.41 ($N = 53$). Former members of the student-centered classes achieved a mean score of 28.84 ($N = 74$).

Inasmuch as the instructor's major goal was to help the students use class subject matter as a guide for future professional behavior, the feelings of the student, i.e., his enthusiasm for the class and the subject matter, would seem to be important. Questionnaire results at the close of the course indicated some significant differences in the perceptions of the students in the instructor-centered and student-centered sections.

(1) Compared to other graduate courses in this general area, I

have learned:

	<i>Much less</i>	<i>Less</i>	<i>About the same</i>	<i>More</i>	<i>Much more</i>	
Instructor-centered	4	21	68	23	15	(N = 121)
Student-centered	2	8	59	53	12	(N = 141)

$$\chi^2 = 21.20 \text{ sig. } < 0.01.$$

(2) As far as learning material which will be of practical use in my life, compared to other courses in this general area, I have learned:

	<i>Much less</i>	<i>Less</i>	<i>About the same</i>	<i>More</i>	<i>Much more</i>	
Instructor-centered	3	14	58	36	13	(N = 124)
Student-centered	0	5	45	68	23	(N = 141)

$$\chi^2 = 20.50 \text{ sig. } < 0.01.$$

(3) Have you had any changes in attitude (not knowledge) as a result of this course?

	<i>Yes</i>	<i>No</i>	
Instructor-centered	49	36	(N = 85)
Student-centered	98	26	(N = 125)

$$\chi^2 = 11.00 \text{ sig. } < 0.01.$$

(4) Compared to other classes, this class has been:

	<i>Very boring</i>	<i>Dull</i>	<i>The same</i>	<i>More interesting</i>	<i>Much more interesting</i>	
Instructor-centered	3	18	60	36	5	(N = 122)
Student-centered	0	12	23	79	31	(N = 145)

$$\chi^2 = 54.00 \text{ sig. } < 0.01.$$

It is quite apparent that from the standpoint of the students, the student-centered courses were, on the whole, better and more interesting learning experiences.

It was not possible to obtain an objective measure of professional changes which could be ascribed to course participation. At the end of the course students estimated the amount of behavioral change that would result from class participation.

(5) How much change in your behavior will result from this course? (Guess)

	<i>None</i>	<i>Little</i>	<i>Pretty much</i>	<i>A great deal</i>	
Instructor-centered	7	70	34	12	(N = 123)
Student-centered	1	55	69	18	(N = 143)

$$\chi^2 = 17.90 \text{ sig. } < 0.01.$$

Once more we have a significant difference in favor of the student-centered classes. Six months after completing the course students were asked if any behavioral changes had actually taken place.

(6) Have there been any changes in your own behavior which you could attribute to the course in "The Psychology of Child Development?"

	Yes	No
Instructor-centered	38	16
Student-centered	68	8

$$\chi^2 = 7.63 \text{ sig. } < 0.01.$$

It can be seen that the earlier perceptions of the students held up pretty well.

The questionnaire also indicated that there was no significant difference between the two groups in the amount of studying done in the area of child psychology or the use of the text as a reference in the six-month period following the class. It is probably safe to assume that graduate courses in education are designed to augment the professional equipment of the students enrolled in such courses. There is, however, a lack of evidence that most courses in education do help the students become better teachers. When students are asked whether they attempt to apply in their own classrooms the theories dealt with in these courses they say, "That stuff is too theoretical. It would never work in our school," or, "I get enthused in class but when I try it out it isn't practical." If these teachers are to actually apply the subject matter dealt with in such courses it is probable that two changes must be made. First, a change in attitude is necessary. Second, the subject matter must be learned more thoroughly.

It is easy to assume that a change in knowledge will result in behavioral changes. Too often this is not the case. If behavioral changes are to occur they must be preceded by changes in attitude as well as knowledge.

If the instructor's goal is an increase of knowledge, the method used to bring about such an increase may well be of little import. If the instructor's goal is to change the professional behavior of his students (so that their classroom performance is guided more by scientific principles and less by animism) the method used may be of major importance.

Take, for example, the use of verbal controls in the classroom. Evidence indicates that the type of controls used by the classroom teacher is correlated with the teacher's effectiveness (10). Informal, unpublished studies suggest that teachers of all ages can make changes in the type of controls they use in the classroom. In each instructor-centered class the instructor included in his lecture a suggestion for a simple method of self-training for teacher improvement in the use of verbal controls. A few weeks later the instructor asked the class if anyone had attempted to use this method of self-training. Approximately two per cent of the students said that they had done so. If the matter of verbal controls was brought up in a student-centered class, the instructor mentioned this same method for self-training. In subsequent classes a student generally brings out the fact that he tried this method. Other students relate their use of the method and the discussion usually indicates that between eight and twelve per cent of the students have actually attempted self-training in the use of their classroom verbal controls. On an examination, students in the instructor-centered classes easily recall that the type of verbal controls used in the classroom is important. They have, then, the knowledge necessary to make an improvement in their teaching. However, they seldom apply this knowledge in their own classrooms. What is needed is not the simple acquisition of knowledge but a change in behavior. This is generally not brought about through reading and listening.

There is evidence that such changes in behavior can be achieved through group discussion and decision. Lewin's data (8) suggest that decisions involving behavior changes are more easily effected by group discussion than by listening to a lecture. A follow-up in one case showed behavior changes as a result of group discussion in thirty-two per cent of the members. The lecture method produced behavior changes in only three per cent of the listeners. Subsequent variations of this experiment followed the same pattern. The change after lectures was in all cases smaller than the change after group discussion. We have, therefore, an indication that the lecture method of teaching, although it may be an efficient vehicle for the transfer of knowledge, may not be an efficient method to produce behavioral change in the student.

It is difficult to estimate the affective state of the students in these thirteen classes. The questionnaire obtained at the final

class meeting indicated that the students generally felt free to participate within both class frameworks.

(7) Did you feel free to make contributions to the class discussion?

	Yes	No
Instructor-centered	78	9 (N = 87)
Student-centered	113	11 (N = 124)

$\chi^2 = 0.11$ no significant difference.

The increased interaction in the student-centered classes encouraged greater friendliness and self-involvement. Many of the one hundred and fifty students enrolled in the instructor-centered classes took no part in group discussions. For some of them their classroom cocoon was undoubtedly comfortable. Without any prodding on the part of the instructor each of the one hundred and forty students in the student-centered classes made some contribution.

No official attendance was taken but more absences were observed in the instructor-centered courses. This may have been a reflection of the higher student interest in the student-centered groups. This interest, in turn, may have been a function of the novelty of this learning situation.

Six of the nine student-centered classes seemed to develop rather high morale. During the closing meetings four of these made an effort to keep the group together through the scheduling of another class in the same community. Two of these groups became so involved in the subject matter that they occasionally extended the class time beyond the closing hour even though the instructor had to leave to teach another class.

SUGGESTIONS FOR THE USE OF A STUDENT-CENTERED METHOD

In a graduate course in education students are usually very curious about the workings of a student-centered method. They elect such a method, however, with a few misgivings. If the instructor limits his contributions to helping the class coordinate the discussion, the first few meetings are likely to go rather well. Students who enjoy talking are given an adequate chance to do so and students who are more reticent are pleased to learn that they will not be forced to participate. After five or six hours of discussion

the more dominant ego needs have been met and a drop in interest and amount of participation may be encountered. It is quite likely that the first few discussions have been interesting only because the people have been interesting. The group has not been very skillful in exploiting the subject matter. It is probably wise for the instructor to prepare the class for this letdown in morale in advance of its occurrence. After four or five hours of discussion have taken place the instructor might make a statement like this. "It seems to me that class has gone rather well. It would not be at all surprising if we had a letdown in interest or morale. If this does occur it probably will not last very long. Let's see it through. The class decided to elect this method of learning and after getting to know you better I feel that it was a wise choice. If, however, midway through the semester you would like to change to another method of learning it will be quite possible to make such a change." Normally, after a few uninteresting meetings the group will grow in discussion skill and will learn to profit academically and emotionally from this new classroom procedure.

Periodical evaluation ought to be very helpful. This can be done through either post meeting reaction forms (probably unsigned) or a general discussion of "how are we doing as a group?" It is probably best to vary evaluatory methods. Sometimes it is helpful for the instructor to ask someone to observe the problem-solving procedures used by the group and to give a short report to the group at the close of the meeting. Critical or even hostile people can be quite useful to the group in this capacity. After such an experience the evaluator often becomes a much more constructive group member.

In a student-centered learning situation the text materials which form the basis for discussion ought to be relatively free of dogmatism. If one text is used it is desirable that it present with some clarity a variety of data and a minimum of value judgments. This encourages students who tend to accept the truth of all printed matter with little or no thought to form their own opinions on the basis of the data and the class discussions.

It seems fairly clear that class size and physical conditions are limiting factors to the usefulness of the student-centered method. In the course of this experiment six of the student-centered classes were seated at a rectangular or square arrangement of tables. Three of the classes were seated in chairs in a circular arrangement.

The instructor felt that these three classes were slower than average in the development of group spirit and a relaxed, free atmosphere. The best possible arrangement would probably be a circular table of such a size that members would be sitting close together but not cramped. The larger the group the more important it becomes that the meeting be held in a room with good acoustics and a minimum of outside noise. Under normal conditions, as the size of the group increases beyond approximately twenty, the forces inhibiting free communication also increase. In an informal discussion group of thirty people, if some members are to contribute they apparently need considerably more motivation to speak than they would need as members of a group of twenty. In the present experiment no classes of more than twenty-five used the student-centered method. This was based on an arbitrary judgment of the experimenter. The data yield no relationship between class size and student performance or satisfaction within either of the learning situations. Under optimum physical conditions, and with a wise, warm and understanding teacher, it is possible that larger classes might profit from such an approach.

One of the difficulties faced by those who wish to experiment with classroom learning methods is the impossibility of duplicating classroom situations. Not only is each class different by virtue of different student personalities, but each instructor's needs are different and the restrictions imposed by the institution are quite dissimilar. Previous research concerning classroom learning methods has largely ignored the ego-needs of the instructor (9, 11). These needs play a vital rôle in the results obtained. On the basis of this research, the experimenter suggests that it is wise for the instructor to be as honest in regard to his ego needs as the situation will allow. In a truly student-centered classroom everyone in the room shares responsibility for the content and the methods used to handle that content. If the instructor feels that he would be derelict in his professional duties if certain aspects of the course content are not adequately dealt with in class, then he ought to express his feelings at the outset. Such a straightforward approach will be appreciated by the students and will interfere less with their self-motivation than a more circuitous method.

The stimuli for instructor self-gratification are quite different in the student-centered classroom. The glow of self-appreciation following the presentation of a well-planned lecture may be sorely

missed. There are, however, multiple satisfactions to the instructor using such a method. Many of these stem from the new insights gained and expressed by students. One member of a student-centered class in a highly rural section made only one contribution during the first five weeks of class. She volunteered the information that she was late because the hired man was sick and she felt that it was more important to "milk those poor cows than to get to class on time." During the sixth class session the topic was "The Affective Life of the Child." Midway through the meeting she said: "I think what we are trying to say is that learning is like a radio. Radio waves are going through the air all of the time but the radio doesn't pick them up until the switch is turned on. Things are happening around us all the time, too, but we don't learn until the switch is turned on. Our emotions are that switch." This was rewarding.

SUMMARY

This study was designed to investigate the effectiveness of two learning situations in achieving the following goals: (1) to help students learn as much as possible about present knowledge in the area of child psychology, and (2) to help students use this knowledge to guide their own behavior.

An instructor-centered approach was used in four sections and a student-centered method was used in nine sections. Ninety per cent of the students in these classes were school teachers, eighty-one per cent were graduate students, and seventy-four per cent were parents. Data were obtained through questionnaires administered at the end of the course and through a test and questionnaire administered six months after the course was completed.

On the basis of four counts—(1) more intervening courses in education and psychology, (2) slightly higher proportion of graduate students, (3) slightly higher mean I.Q., (4) familiarity with test style and test items—it might be expected that members of the instructor-centered classes would achieve higher scores on the test. Actually no significant difference was found. Students who had participated in the instructor-centered classes achieved a mean raw score of 28.41 ($N = 53$). Former students in the student-centered classes achieved a mean raw score of 28.84 ($N = 74$).

At the close of the course, members of the student-centered sections estimated that they had learned more, that what they

had learned would be of more practical use to them, that more attitude change had taken place, that more behavioral change would take place as a result of the course, and that the class had been more interesting. All differences between the instructor-centered and student-centered groups were significant at greater than the 0.01 level and favored the student-centered sections.

During the six-month period following the class there was no significant difference between the two groups in the amount of studying done in the area of child psychology or the use of the text as a reference. Six months after the completion of the course, the student-centered sections estimated more behavior change as a result of the course than did the instructor-centered sections (significant at better than 0.01).

REFERENCES

- 1) M. J. Asch, "Nondirective Teaching in Psychology; An Experimental Study," *Psychology Monograph*, 65: 4, 1951.
- 2) E. W. Bovard, Jr., "The Psychology of Classroom Interaction," *Journal of Educational Research*, 45: 215-224, 1951.
- 3) E. W. Bovard, Jr., "Clinical Insight as a Function of Group Process," *Journal of Abnormal and Social Psychology*, 47: 534-539, 1952.
- 4) N. Kantor, *The Dynamics of Learning*. Buffalo, Foster and Stewart, 1946.
- 5) V. Faw, "A Psychotherapeutic Method of Teaching Psychology," *American Psychologist*, 4: 104-109, 1949.
- 6) L. M. Gibb and J. R. Gibb, "The Effects of the Use of 'Participative Action' Groups in a Course in General Psychology," *American Psychologist*, 7: 247, (abstract), 1952.
- 7) H. Kelley and A. Pepitone, "An Evaluation of a College Course in Human Relations," *Journal of Educational Psychology*, 43: 193-209, 1952.
- 8) Kurt Lewin, "Group Decision and Social Change," *Readings in Social Psychology*, pp. 459-473. New York, Henry Holt and Co., 1952.
- 9) W. J. McKeachie, "Student-Centered versus Instructor-Centered Instruction," *Journal of Educational Research*, 45: 143-150, 1954.
- 10) W. C. Olson and M. M. Wilkenson, "Teacher Personality as Revealed by the Amount and Kind of Verbal Direction Used in Behavior Control," *Educational Administration and Supervision*, 24: 31-39, 1938.
- 11) H. Ruja, "Experimenting with Discussion in College Teaching: a Survey of Recent Research," *Educational Administration and Supervision*, 39: 321-342, 1953.

THEORY DEVELOPMENT AND THE STUDY OF TEACHER BEHAVIOR

DAVID G. RYANS

University of California at Los Angeles

THE RÔLE OF THEORY

Teacher behaviors, their nature, genesis, and cultivation, their identification and their evaluation, constitute the core concern not only of teacher training institutions and school systems, but also of a society at large that depends on teachers to a very great extent for the propagation of accumulated knowledge and cultural values. Considerable opinion is expressed and not a few studies are undertaken in an effort to better understand teacher behavior. (1, 2, 4, 5, 8, 10, 11, 13, 14.) Relatively little attention, however, appears to have been directed toward the organization or systematization of what is known about teacher behavior or the assumptions on which the study of teacher behavior is based.

In the meantime, strong arguments have been presented by social and behavioral scientists suggesting that steps toward some sort of a theoretical formulation have considerable utility value; that systematic theories are highly desirable, particularly for guidance, if maximum productivity of research in a field is to be attained. Proponents of this view point out that advanced understanding and usable knowledge frequently have been observed to demonstrate marked increase as the study of problems in an area progressed from exploratory, "catch-as-catch-can," investigations to selective observation guided by hypotheses derived from systematic theory and employing empirical tests to determine the place (if any) of such hypotheses in the basic theory.

Essentially, this approach represents a rapprochement of formal, deductive logic and a modern science that demands the inclusion of pragmatic considerations in judging the admissibility of facts to a body of knowledge. Such axiomatic, or hypothetico-deductive, theory is rather popular at the present time.

Another sort of theorizing attempts to be rigorously inductive and culminates in what may be simply described as an organized body of empirically obtained facts. This kind of theory employs operational definitions extensively and emphasizes reasoning that

proceeds from the observed characteristics of sample data to generalizations applicable to larger homogeneous classes. Facts are admissible to such a theory if (1) they are based on fair samples, (2) they satisfy pragmatic criteria, and (3) they fit into a pattern with other known facts in the same area. Hypotheses may be derived from the pattern of accumulated fact (and to this extent the approach is also deductive). Inductively generated theories, or systematizations, are descriptions rather than explanations.

Researchers and practitioners appear to have found little need for either kind of theory development in approaching the problem of teacher behavior. A recent statement attributed the lack of productive research on teacher effectiveness, in part, to such a neglect of theory, declaring:

... the present condition of research of teacher effectiveness holds little promise of yielding results commensurate with the needs of American education. This condition has two significant characteristics: disorganization, and lack of orientation to other behavioral sciences. By disorganization, we mean the condition in which, at present, research too often proceeds without explicit theoretical framework, in intellectual disarray, to the testing of myriads of arbitrary, unrationalized hypotheses. The studies too often interact little with each other, do not fall into place within any scheme, and hence add little to the understanding of the teaching process. The simple fact of the matter is that, after 40 years of research on teacher effectiveness during which a vast number of studies have been carried out, one can point to few outcomes that a superintendent of schools can safely employ in hiring a teacher or granting him tenure, that an agency can employ in certifying teachers, or that a teacher-education faculty can employ in planning or improving teacher-education program. (2: p. 657.)

Something more should be said, perhaps, of the essential characteristics of theories and the values that are claimed for them. Rose, in his volume of essays on theory in the social sciences, states rather well the generally accepted concept of theory viewed from the standpoint of formal logic with appropriate adaptations to make place for a pragmatic criterion of validity. Rose writes that theory is:

... an integrated body of definitions, assumptions, and general propositions covering a given subject matter from which a comprehensive and consistent set of specific and testable hypotheses can be deduced logically. The hypotheses must take the form "If a , then b , holding constant c, d, e, \dots " or some equivalent of this form, and thus permit of causal explanation and prediction. (12: p. 3.)

TABLE I.—THEORETICAL FRAMEWORK

Type I. Axiomatic, Mathematico-Deductive, Hypothetico-Deductive Theory	Type II. Inductive, Empirical Theory
<p>A. Assumptions Definitions</p> <p>B. General Propositions</p> <p>C. Hypotheses: postulates and theorems re properties and functional relationships</p> <p>D. Observation: testing of postulates and theorems by classification and experimentation</p> <p>E. Extension and revision of theory in light of observation and of further deductive analysis</p> <p>F. Prediction</p>	<p>A. Assumptions Definitions (especially "operational")</p> <p>B. Observation Classification and cross-classification Sporadic hypotheses and their experimental testing</p> <p>C. Specified descriptions of properties and functional relationships (generalizable to inductive principles and laws)</p> <p>D. Systematization, organization of descriptive principles into coherent pattern (possibly leading to General Propositions)</p> <p>E. Extension and revision of principles, and of theory, as required by further classification and experiment</p> <p>F. Prediction</p>

form his teaching responsibilities adequately." "I think older teachers are more authoritarian than younger ones."). But all such ruminations probably should not be thought of as hypotheses—at least, not in the sense in which the term hypothesis usually is employed in science. There are certain generally agreed-upon restrictions that commonly are applied. Thus, a suitable hypothesis must be in an exact form; must provide a specific answer to a specific question. And it must not be inconsistent with what is known; it must not be contradictory to the basic assumptions and propositions of applicable theory, or with other hypotheses for which there is extensive supporting evidence. And, again, a suitable hypothesis must be stated in a way that it is capable of being confirmed or refuted at some agreed-upon level of confidence; the consequences that will follow if it is probably true, or probably false, must be describable and such consequences must be capable of observation and recording.

SOME STARTING POINTS FOR A THEORY
OF TEACHER BEHAVIOR

Thus far our discussion has been concerned with comments on the nature of theory. Now how can some of these concepts relating to theoretical framework be applied to teacher behavior? What, for example, might be some of the assumptions and definitions that apply to teacher behavior? Those that will be suggested do not constitute a complete set of all assumptions required for a theory of teacher behavior. Nor is any particular claim made at this point for theoretical rigor. But if we do wish to think in terms of theory, a starting point is necessary regardless of how crude and tentative such a beginning may be. Trial and error, criticism, and research may provide the necessary clarification, revision, and extension leading to a generally acceptable theory of teacher behavior.

A definition. As a point of embarkation, a definition of teacher behavior will be stated. It is suggested that for our purposes teacher behavior may be defined simply as the behavior, or activities, of persons as they go about doing whatever is required of teachers, particularly those activities that are concerned with the guidance or direction of the learning of others.

There are several important implications in this rather straightforward operational definition.

(A) One implication is that teacher behavior is social behavior—that, in addition to the teacher, there must be learners, or pupils, who presumably are affected by the behavior of the teacher. It should also be noted that teacher behavior-pupil behavior relationships may be of a reciprocal nature; that not only do teachers affect pupil behavior, but pupils may affect teacher behavior, as well.

(B) Another implication of the definition of teacher behavior is that there is nothing inherently good or bad in a given teacher behavior, or set of behaviors, but, instead, that teacher behavior is good or bad, right or wrong, effective or ineffective, only to the extent that the behavior conforms or fails to conform to the particular value system, or set of objectives, defining (1) the activities expected of a teacher in a given community or culture, and (2) particularly the kinds of learnings (attainments) and methods of teaching to bring about these learnings, approved by the particular culture (11, 14).

Assumption I. If we were to state the basic assumption to be

made in setting out to describe teacher behavior, we might expect it to bear some resemblance to formulations made for similar purposes in connection with learning theory and personality theory. Indeed, in any and all behavior theory, some expression of faith in the reliability, or consistency of behavior is required. In our case the assumption may be summarized in the proposition that teacher behavior is a resultant of (a) certain situational factors and (b) certain organismic factors, and their interaction; or, simply that teacher behavior is a function of certain environmental influences and characteristics of the individual teacher. Thus:

$$(A) \quad r_a = f[(S_{1a}M_1^{\circ}E_1^{\circ}H_1^{\circ}) + (S_{2a}M_2^{\circ}E_2^{\circ}H_2^{\circ}) + \dots (S_{na}M_n^{\circ}E_n^{\circ}H_n^{\circ})]$$

where:

r_a = teacher behavior r in situation a

S = situational indices

M° = motivational organismic indices

E° = experience organismic indices

H° = genetic organismic indices

(B)

$$R_t = ar_a + br_b \dots nr_n + e$$

where:

R_t = over-all teacher behavior

a = weight

r_n = z score of teacher behavior n

e = z score of error component

The present treatment is intentionally limited to consideration of teacher behavior per se. Since, however, teacher behavior implies pupil behavior (as previously noted), we might also observe that pupil behavior could be thought of as being similarly influenced by a number of variables, one of which would be teacher behavior R_t . Thus:

(C)

$$r_a + f[(S_{1a}M_1^{\circ}E_1^{\circ}H_1^{\circ}) + \dots (S_{na}M_n^{\circ}E_n^{\circ}H_n^{\circ})]$$

where:

S represents situational indices such as textbooks, peer behavior, parent behavior, teacher behavior, etc. and

(D)

$$R_p = ar_a + br_b \dots nr_n + e$$

where: R_p = pupil behavior

What these rather rigid appearing equations say is simply that behavior occurs when an organism is stimulated and if the variables and mathematical functions are known, the equation may be solved for the unknown (r or R).

Just how these various situational and organismic conditions interact and what takes place in the teacher's nervous system as they interact (how, as some theorists would put it, the energy input-output transfer takes place) certainly is not known and we are completely incapable of describing the process except in terms of inferences based upon observable¹ inputs and observable responses of the teacher. The fact that little is known about how such things take place does not mean that persons interested in behavior theory have not been actively concerned about the problem. Certain groups of theorists have been both active and ingenious. One such group, which incidentally is interested in speculating upon the generality of behavior theory whether the systems involved are atoms, viruses, cells, individual persons, society, solar systems, or what not, view the organism, the teacher in our case, as an open system (i.e., a bounded region in space-time), possessed with negative feedback which provides for the distribution of information to sub-systems to keep them in orderly balance.

Other theorists have been particularly concerned with how individuals in a social environment (e.g., teacher-pupil, teacher-teacher, etc.) interact and condition one-another's behavior, not only in an immediate situation but also in future situations, as a result of the integration of response-produced stimuli into the total stimulus pattern. Sears (15) has suggested, for example, the necessary expansion of the basic monadic unit of behavior, which various learning theories have employed—and in terms of which the foregoing equations (A), (B), (C), and (D) are presented—into a diadic one which describes the combined actions of two or more individuals. That a diadic unit is essential if the relationships between people are to be taken into account in theory should impress the teacher behavior theorist as an entirely reasonable proposition.

The diadic approach strikes at the heart of the teacher-pupil relationship problem. Therefore, it is important to take into account extension of the paradigm represented by our equation (A)

¹ Observable, as here used, refers to that which may be perceived either by another person or by the experiencing individual, or which may be recorded by some instrument.

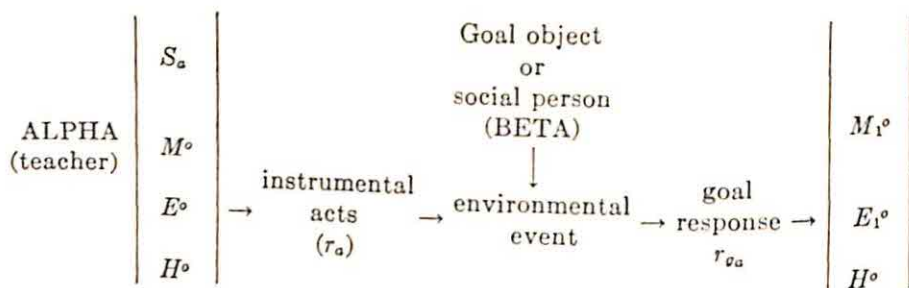


FIGURE 1. The monadic instigation-action sequence. After Sears (15)

to include the significant implications of Sears' model for the diadic behavior sequence. The models which follow conceptualize teacher behavior first in terms of the monadic and then the diadic behavior sequence. Alpha may be presumed to be the teacher, and Beta the pupil. The symbols S_a , M^o , E^o , H^o , and r_a have the same meaning as in the earlier equations. M_1^o and E_1^o are employed to represent the modified motivational and experience structures of the teacher resulting from a reinforcement brought about by completion of the behavior sequence.

We are proposing, then, that the basic assumption of teacher behavior theory may be summarized by the equations and the models just presented.

Now growing out of the basic assumption that teacher behavior is a function of the conditions under which it occurs are a number of implications or sub-assumptions, which follow:

(A) An implication of this assumption is, of course, that teacher behavior (and social behavior, with which education deals), is characterized by some uniformity; that, as Mill put it: "... there

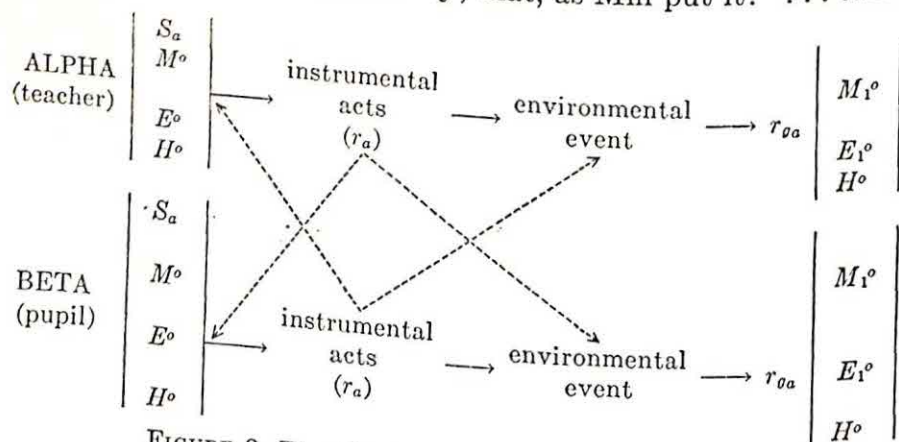


FIGURE 2. The diadic sequence. After Sears (15)

are such things in nature as parallel cases, that what happens once will, under sufficient degree of similarity of circumstances, happen again" (9: p. 223). We are stating that teacher behavior is, within limits, reliable, consistent, or stable.

(B) Another implication of this basic assumption (perhaps it is so fundamental to any or all scientific theory that it is unnecessary to state it explicitly with respect to teacher behavior), is expressed by Keynes' Postulate of Limited Independent Qualities, which states that: "... objects in a field over which our generalizations extend, do not have an infinite number of independent qualities; ... their characteristics, however numerous, cohere together in groups of invariable connections, which are finite in number" (7: p. 256). This is to say that the number of responses the individual teacher is capable of making, and the number of stimulus situations and organismic variables that may affect a teacher's behavior, are limited. The assumption is an important one if the equations noted above are to hold, and, of course, if we hope to predict teacher behavior. It presents the researcher with a "tolerable" problem (as a student recently put it).

(C) In taking cognizance of this basic proposition of teacher behavior, it should be noted that human behavior, characterized as it is by variability (only approximate uniformity, or consistency), always must be considered in light of probability rather than rigid, one-to-one, invariable cause-effect relationships. The error component (unreliability) resulting from behavior sampling inevitably will be present in any assessment we make of stimulus conditions, of organismic conditions (genetic bases, past experience, motivation), and of the teacher behavior (the dependent variable, or criterion). Teacher behavior can be predicted only with varying degrees of probability; never with certainty. Therefore, r_a and R_t refer to probable behaviors rather than certain behaviors.

Assumption II. When we presume to study teacher behavior we also make the assumption that teacher behavior is observable; that its manifestations are of a tangible nature and may be identified objectively, either by direct observation of samples of behavior, or by approaches that provide correlative indices of teacher behaviors (e.g., approaches involving the assessment of pupil behavior, the use of tests of teacher abilities and knowledges, the use of interviews or inventories to elicit expression of teacher preferences, interests, beliefs, and attitudes, etc.).

Several implications (or perhaps subassumptions) may be noted here.

(A) One implication is that different teacher behaviors, or sets of teacher behaviors, are distinguishable from one another; that some behaviors possess characteristics in common, to the exclusion of other characteristics. Teacher behaviors can be distinguished under observation.

(B) A second subassumption is that the distinguishing characteristics of a teacher behavior provide the basis for a concept, abstracted from specific instances, which permits common understanding of the behavior and communication of the behavior to others.

(C) These implications lead to a third aspect of the assumption of the observability of teacher behavior, that teacher behaviors are classifiable, both qualitatively and quantitatively. (1) Teacher behaviors that are similar, that have common elements, may be classified in the same category—apart from other teacher behaviors that are dissimilar. (2) Within any given category teacher behaviors may be further assigned to subclasses that may be treated quantitatively (to which numbers may be attached). These quantitative subclasses may be of a present-absence type permitting enumeration, or counting, only; or, they may be subclasses characterized by continuity and varying as a metric (exemplified by ordinal subclasses at the lowest level of refinement and by equal-interval and equal-ratio subclasses at successively more refined levels).

Various qualitative classifications of teacher behaviors are possible. Thus, for example, we might think of teacher behaviors grouped broadly into such general categories as: (1) those involving instruction and relationships with pupils; (2) those involving relationships with the school, its organization, and its administration; and (3) those involving relationships with the community. Or, we might classify teacher behaviors more specifically, employing such categories as verbal aptitude, emotional stability, favorable attitude toward pupils, friendly behavior toward pupils, and the like, to provide the framework. For any specifically defined category we might conceivably assign quantitative subclasses.

(D) Still a fourth subassumption to the basic assumption of observability of teacher behavior might state that teacher behaviors may be revealed, or may be observed, either (1) by the representative sampling of teacher behaviors or (2) by signs, or

indicators, or correlates of the behavior under consideration. In sampling behavior, we assume that the performance of the individual during the behavior sample is proportional to the larger aspects, or universe, of his behavior. In judging behavior from signs or correlates, it is assumed that behavior can be inferred from known correlates of that behavior—from phenomena that have been found to accompany that behavior in the past.

So much for our articles of faith, or general assumptions regarding teacher behavior. With such assumptions stated (and assuming for the present the desirability of inductive theorizing), the teacher behavior theorist might turn to observation—to classification and hypothesis testing.

The number of descriptive classifications and specific hypotheses that might be generated with respect to teacher behavior is almost limitless. Probably no one would be interested in all such hypotheses even if it were possible to assemble them. But some classifications and some hypotheses are more relevant than others. And, moreover, some appear to be quite plausible. In fact, many can even be incorporated in the design of well-controlled study.

At this point it might be well to note that there are various approaches to the testing of hypotheses. Ideally, to test any hypothesis, we would like to design a projected experiment (the basic experimental design) with (1) experimental and control groups equated at the beginning of the investigation, (2) introduction of an experimental variable or treatment (the hypothesized causal factor) with the experimental group while withholding it from the control group, and (3) judgment of the validity of the hypothesis by comparison of the groups with respect to the dependent variable, or criterion behavior. This is the typical application (or, it may involve some more advanced and complex modification to permit analysis of interaction effects and such), of Mill's methods of agreement and difference. But this approach frequently is not possible in the study of teacher behavior. Instead, it usually is necessary to undertake what Chapin (3) has called "ex post facto" experiments, or perhaps the comparison of aspects of defined groups, or perhaps intercorrelational studies (classificatory investigation). Much of the research on teacher behavior must be of this nonexperimental (using "experiment" in the strictest sense) type.

No attempt will be made to state specific hypotheses about

teacher behavior that might be subjected to testing. Tests of a number of hypotheses have been attempted in connection with various researches (10), including the Teacher Characteristics Study (14). In the latter research, hypotheses referring to the interrelationships of estimates of teacher behaviors have been extensively tested. Similarly, hypotheses relative to the central tendencies and dispersions of behavior measurements of defined groups of teachers have been given attention, leading to comparisons of teachers with respect to such conditions as age, experience, geographic area of employment, type of college attended, youthful experiences, frequent activities, etc., on a selected set of teacher personal and social characteristics.

Throughout the country, many similar researches are being directed at the study of comparisons of, and interrelationships among, teacher behaviors. Perhaps we are reaching the point where patterns among the behavior components may be discerned—where the beginnings of organization of evidence into a theory of teacher behavior may be possible.

Such study and systematization (theory development), in combination with proper attention to specification of problems and careful design, may reasonably be expected to shed increasing light upon our understanding, first, of teacher behavior, and, second (and of ultimate importance), on how teacher behavior influences pupil behavior and how certain kinds of teacher-pupil relationships may be identified, predicted, and cultivated.

REFERENCES²

- 1) A. S. Barr, B. V. Bechdolt, W. W. Cox, N. L. Gage, J. S. Orleans, H. H. Remmers, and D. G. Ryans, "Report of the Committee on the Criteria of Teacher Effectiveness," *Review of Educational Research*, 22: 238-253, 1952.
- 2) A. S. Barr, B. V. Bechdolt, N. L. Gage, J. S. Orleans, C. R. Page, H. H. Remmers, and D. G. Ryans, "Second Report of the Committee on Criteria of Teacher Effectiveness," *Journal of Educational Research*, 46: 641-658, 1953.
- 3) F. S. Chapin, *Experimental Designs in Sociological Research*. New York, Harper & Bros., 1947.

² References 1, 2, 11, and 14 bear, to some extent, on theory of teacher behavior. References 5 and 10, in particular, provide bibliographies and reviews of studies relating to the evaluation of teacher behavior. It should be noted, also, that the June issues of the *Review of Educational Research* in the years 1934, 1937, 1940, 1943, 1946, 1949, 1952, and 1955 were devoted to "Teacher Personnel."

- 4) W. W. Charters and D. Waples, *The Commonwealth Teacher Training Study*. Chicago, University of Chicago Press, 1929.
- 5) S. J. Domas and D. V. Tiedeman, "Teacher Competence: An Annotated Bibliography," *Journal of Experimental Education*, 19: 101-218, 1950.
- 6) E. Greenwood, *Experimental Sociology*. New York, King's Crown Press, 1945.
- 7) J. M. Keynes, *A Treatise on Probability*. New York, Macmillan Co., 1921.
- 8) W. A. McCall, *Measurement of Teacher Merit*. Raleigh, N. C., State Department of Public Instruction, 1952.
- 9) J. S. Mill, *A System of Logic*. New York, Longmans, Green, Co., 1872.
- 10) J. E. Morsh, and E. W. Wilder, *Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1952*. San Antonio, Texas, Air Force Personnel and Training Research Center, 1954.
- 11) W. Rabinowitz and R. M. W. Travers, "Problems of Defining and Assessing Teacher Effectiveness," *Educational Theory*, 3: 212-219, 1953.
- 12) A. M. Rose, *Theory and Method in the Social Sciences*. Minneapolis, University of Minnesota Press, 1954.
- 13) D. G. Ryans, "Appraising Teacher Personnel," *Journal of Experimental Education*, 16: 1-30, 1947.
- 14) D. G. Ryans, "The Investigation of Teacher Characteristics," *The Educational Record*, 34: 370-396, 1953.
- 15) R. R. Sears, "A Theoretical Framework for Personality and Social Behavior," *The American Psychologist*, 6: 476-483, 1951.

RESULTS AND DISCUSSION

The null hypotheses to be tested by the application of analysis of variance to the data are: (1) there is no significant difference between the sexes; (2) there is no significant difference between the two social classes; and (3) the interaction between sex and social class status is not significant.

In Table I are given the means and standard deviations for the two social classes and for the sexes separately as well as for the total group, by sex and by social class. The analysis of variance results are presented in Table II.

The evidence indicates that there are no significant differences between the sexes or between the classes for any of the three measures of verbal facility used. The interaction between sex and class is likewise not significant. Although statistically the differences are not significant, the means of the three measures as observed in Table I reveal a consistent trend. The boys are superior to the girls and maintain their superiority in all three measures, regardless of social class position. On the other hand, when sex is disregarded, the high social class children are always superior to the low social class children. In all three verbal measures, the girls of the low social class are most disadvantaged.

It is possible that significant differences were not obtained because of the large individual variation present (note S.D.'s and error variance). The most nearly significant difference (between the two social classes in the number of uncommon words used) deserves special mention. Although very bright children seem able, to some

TABLE I.—MEANS AND STANDARD DEVIATIONS BY SEX AND SOCIAL CLASS FOR THREE MEASURES OF VERBAL FACILITY

	All Boys	All Girls	All Low Class	All High Class	Low Class Boys	High Class Boys	Low Class Girls	High Class Girls
Total number of words used	55.7* (26.1)	47.9 (23.3)	47.9 (26.3)	55.7 (24.1)	54.6 (27.8)	56.8 (24.2)	41.3 (22.9)	54.6 (21.9)
Number of uncommon words	2.8 (2.6)	2.4 (2.2)	1.95 (1.8)	3.25 (2.8)	2.0 (1.7)	3.6 (3.0)	1.9 (1.8)	2.9 (2.4)
Number of adjectives and adverbs	5.8 (4.5)	4.3 (3.4)	4.4 (3.8)	5.6 (4.3)	4.8 (4.3)	6.8 (4.5)	4.0 (3.1)	4.5 (3.6)

* The upper number is the mean; the number in parentheses is the standard deviation.

TABLE II.—ANALYSIS OF VARIANCE BETWEEN BOYS AND GIRLS
AND BETWEEN LOW AND HIGH SOCIAL CLASSES FOR
THE THREE MEASURES OF VERBAL FACILITY

	Source	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Total number words used	Sex	600.63	1	600.63	0.91
	Class	600.63	1	600.63	0.91
	Sex x Class	308.03	1	308.03	0.47
	Error	23600.50	36	655.57	—
	Total	25109.77	39	—	—
Number of uncommon words	Sex	1.60	1	1.60	0.27
	Class	16.90	1	16.90	2.84
	Sex x Class	0.90	1	0.90	0.15
	Error	214.20	36	5.95	—
	Total	233.60	39	—	—
Number of adjectives and adverbs	Sex	24.03	1	24.03	1.39
	Class	15.63	1	15.63	0.91
	Sex x Class	5.63	1	5.63	0.33
	Error	619.70	36	17.21	—
	Total	664.97	39	—	—

extent, to overcome the handicaps in language ability resulting from their underprivileged environments, apparently they do not acquire the same facility with difficult and unusual words which their more fortunate peers have attained. This difference is the more striking since the poor children in this study are on the average slightly brighter than the well-to-do children (Mean I.Q.'s 141 and 137).

The finding that boys are superior to girls is contrary to the findings of the majority of previous investigations, and may in part be explained by the fact that the boys were superior to the girls in intelligence, as indicated by differences in average I.Q.'s (143 and 135).

The major implication of this study rests in the obvious conclusion that schools must enrich the experiences of underprivileged children, whether they be dull or bright. The enrichment program should be concerned with verbal experiences and language skills of all kinds.

REFERENCES

- 1) Frank Auld, Jr., "Influence of Social Class on Personality Test Responses," *Psychological Bulletin*, 49: 318-332, 1952.
- 2) Helen H. Davidson, *Personality and Economic Background*. New York, King's Crown Press, 1943.
- 3) Allison Davis, *Social Class Influences Upon Learning*. Cambridge, Harvard University Press, 1948.
- 4) M. I. Dundson and J. A. F. Roberts, "The Relation of the Terman-Merrill Vocabulary Test to Mental Age in a Sample of English Children," *British Journal of Statistical Psychology*, 6: 61-70, 1953.
- 5) Erwin J. Lotsof, "Intelligence, Verbal Fluency, and the Rorschach Test," *Journal of Consulting Psychology*, 17: 21-24, 1953.
- 6) Kurt B. Mayer, *Class and Society*. Garden City, N. Y., Doubleday and Co., 1955.
- 7) Esther Milner, "Effects of Sex Role and Social Status on the Early Adolescent Personality," *Genetic Psychology Monographs*, 40: 231-325, 1949.
- 8) Esther Milner, "A Study of the Relationship between Reading Readiness in Grade One School Children and Patterns of Parent-Child Interaction," *Child Development*, 22: 95-112, 1951.
- 9) Dorothea McCarthy, "Some Possible Explanations of Sex Differences in Language Development and Disorders," *Journal of Psychology*, 35: 155-160, 1953.
- 10) Dorothea McCarthy, "Language Development in Children," in *Manual of Child Psychology*, edited by L. Carmichael, second edition. New York, Wiley and Sons, 1954.
- 11) Meyer Rabban, "Sex-role Identification in Young Children in Two Diverse Social Groups," *Genetic Psychology Monographs*, 42: 85-158, 1950.
- 12) Lewis Terman, "Psychological Sex Differences," in *Manual of Child Psychology*, edited by L. Carmichael, second edition. New York, Wiley and Sons, 1954.
- 13) Florene M. Young, "An Analysis of Certain Variables in a Developmental Study of Language," *Genetic Psychology Monographs*, 23: 1-141, 1941.

AN EXPERIMENTAL COMPARISON OF NON-SENSE, WORD, AND SENTENCE MATERIALS IN EARLY TYPING TRAINING¹

LEONARD J. WEST

Air Force Personnel and Training Research Center
Chanute Air Force Base, Illinois

Commercial typing textbooks as well as practice materials locally prepared by certain military training installations use a variety of materials in the early keyboard learning stages of typewriting instruction. These materials consist either of: (a) isolated words, (b) ordinary sentences, (c) "location drill" sequences made up of nonmeaningful series of letters struck by individual fingers—or of various combinations of these three types of materials.

Materials of the "location drill" or "nonsense drill" type (as they are characteristically called by typing teachers) were in use even before the pioneer experimental work of Book in 1908, republished subsequently together with a textbook for teachers (1, 2). Their continued use on a very much reduced scale in today's commercial typewriting texts was somewhat supported by Book's description of typewriting skill as consisting of mastery over a hierarchy of stroking habits in which the typist first learns individual-letter habits before he can handle larger units of material in such a way that "the various movements required to write certain words and phrases become so closely linked together that the first movement in the series needs only to be purposely initiated and consciously controlled to set off all the movements in the series" (1, p. 173). Book's alternative suggestion that lower-order stroking habits might be perfected incidentally through the use of word materials or of connected discourse was reflected in a dozen attempts (during the period 1921–1948) to evaluate experimentally the merits of nonsense drill materials as against either

¹ This report is based on work done under ARDC Project No. 7714, Task No. 77260, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

word or sentence materials. A small but general tendency toward inferior performance on the part of those trained on nonsense materials in these experiments must be considered somewhat inconclusive largely because of failure to hold constant (a) the rate of introduction of new keys, (b) the frequency of appearance of each letter in the practice materials, and (c) the total amount of practice on each type of material.

The present experiment was designed to hold constant the three factors mentioned in comparing the merits of nonsense, word, and sentence materials as vehicles for the initial learning of key locations. In order to assess the possible influence of word length as a factor in difficulty, one set of isolated word materials was restricted almost entirely to those of three to five letters in length, while another set included words covering a wider range of length (two to seven letters) and averaged 0.5 letters per word more than those of the first set. With a view toward evaluating the four types of materials as preparation for ordinary sentence typing as well as for the typing of the random assortments of letters contained in the five-letter code groups typed by radio operators (e.g., *XIPKW LNUPM*), the final speed and accuracy of the subjects in this experiment were determined both for sentence and for code copy. While this experiment was not designed to be a direct test of any particular hypothesis underlying the nature of the practice materials and must be viewed as purely an empirical comparison of materials, certain factors which may account for the findings are discussed after the presentation of results.

METHOD

Materials. Samples of the four types of materials are shown in Figure 1 together with the practice schedule.

Several features of these materials should be mentioned. First, the nonsense sequences are *not* random assortments of letters, but fixed and rigid sequences specifically used for many years and designed to train each finger in turn to strike the keys assigned to it (e.g., only the left index finger is used for the sequence *frf*). Secondly, each line (or two) of the materials—and the ones shown are only samples—contains every letter of the alphabet. (The inclusion in the nonsense materials of certain characters and punctuation marks is demanded by the nature of the sequences. These are superfluous and were not included in the test materials.)

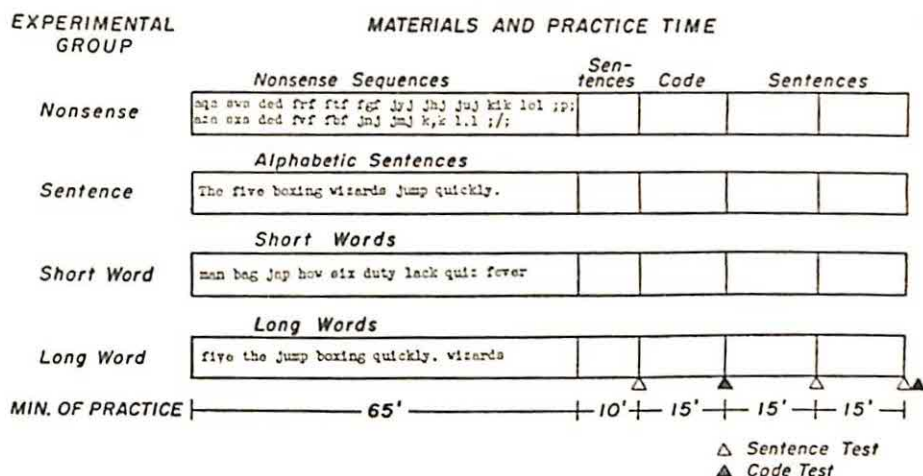


FIGURE 1. Minutes of typing practice on various materials by each of four experimental groups.

Third, it may be noted that the long word materials are simply the sentence materials in scrambled, nonsense order. This was designed to furnish at least one comparison between materials in which letter sequences, in addition to other frequency variables, were held constant. Fourth, on the basis of the entire set of practice materials, those for the three word and sentence sets had identical letter frequencies. Discrepancies for the nonsense sequence typists were more than made up for by the anticipated finding that, because of the ease with which nonsense sequences may be typed, two and one-half times as much typing was done of the nonsense sequences as of any of the other three sets of materials.

Subjects. Three hundred forty-five airmen with no previous typing experience, either formal or informal, were used. Three-quarters of these men were awaiting entry into courses which include typing training. They were assigned to groups (one for each of the four types of materials) at random from within strata based on the course they were about to enter and on their Clerical Aptitude Index—the latter a measure used by the Air Force to predict success in clerical training and used as one basis of assignment to training. The nonsense, sentence, short word, and long word groups had, respectively, eighty-four, eighty, eighty-nine, and ninety-two men.

Procedure. The main procedural features of this experiment are shown in Figure 1. As shown, the first sixty-five minutes of a total

of one hundred and twenty minutes of practice were devoted to the differential materials. The last fifty-five minutes of practice used the same materials for all groups: ten minutes on a sentence designed to introduce capitalization to all learners; fifteen minutes on random, five-letter, equal-frequency code groups; and thirty minutes on sentences containing every letter of the alphabet. The same code and sentence materials were used by all groups in the second stage of practice in order to make the training more nearly correspond to standard procedures in which nonsense drill or word practice is envisioned purely as introductory practice to be followed by the use of materials gradually approaching those used on the job. Criterion measures consisted of speed and error scores on either one or two two-minute timings on new code or new alphabetic sentence materials.

Instructional sessions took place on four consecutive mornings for each of three weeks at the same time each day, in the same room, with the same equipment, and with the same instructor. A different group of men was used for each of the twelve sessions; there were, thus, three "subgroups" for each type of materials. Materials were assigned to groups at random within each four-day set of sessions, each session being devoted to a different one of the four treatments. Tape recordings of all instruction revealed virtually verbatim correspondence of instructor activities and comments throughout.

Each session lasted four hours: two hours of practice, sixteen minutes of testing, the remaining time being devoted to periodic rest breaks and to orientation and verbal and visual instructions. Substantial fatigue was noted during the latter stages of this relatively massed practice, terminal performance thereby being adversely affected.

Other relevant instructional aspects were as follows: (a) Except for the first ten minutes of practice, there was no repetitious practice on individual sequences, words, or even lines; units of six to ten different lines were typed before repetition was allowed. (b) During the initial stages of practice, subjects were urged to look at the keyboard as an aid to locating keys. Increasing emphasis on typing purely by touch was introduced after half an hour of practice. Various mixtures of sight and touch typing were observed to be in use by many learners through the first three-quarters of practice. Thereafter, and for the terminal measures, the mode of

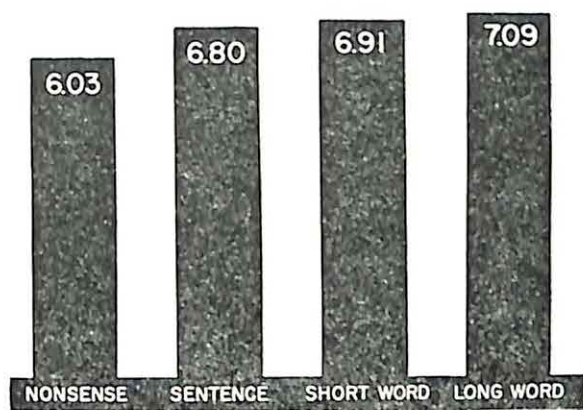


FIGURE 2. Mean words per minute of each of four groups on final, four-minute sentence test.

operation was observed to conform to standard touch typing procedures. (c) For the first line (or two) of practice (five to ten minutes) subjects were led through the materials step by step with explicit directions for each motion. All practice thereafter was at the learner's own rate.

RESULTS

Criterion measures consisted of total strokes (converted here to word-per-minute speeds) and number of errors on two two-minute final sentence tests and one two-minute final test on code materials.² In evaluating differences among and between groups, 0.05 was set as the minimum reliability level. Kruskal and Wallis' nonparametric analysis of variance by ranks (3, pp. 436-440) resulted in two chi-squares with probabilities indicating reliable differences among treatment ranks, namely, (a) 0.001 for sentence speed, and (b) 0.04 for code speed. That for sentence errors was

² Because the distributions of scores did not meet the assumptions concerning homogeneity of variance and normality of distribution desirable for analysis of variance, and because the effect of these discrepancies on analysis of variance was unknown, the data were subjected both to classical analysis of variance and to the nonparametric analysis of variance by ranks. The two types of analysis resulted in virtually identical probability levels. Over-all analyses were followed by *t* tests and by the nonparametric Mann-Whitney "U" test between various pairs of treatments—again producing virtually identical probability levels. In checking on consistency of treatment of the groups from session to session, it was found that there was no significant interaction between sessions and treatments.

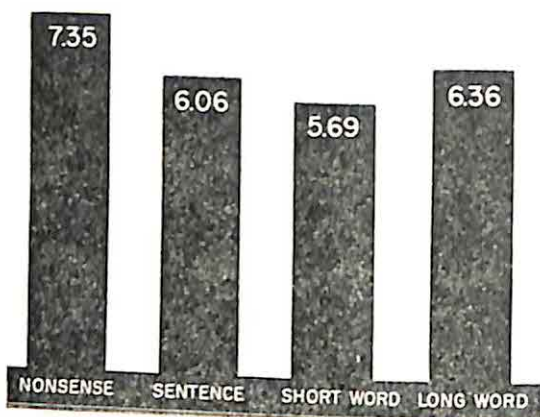


FIGURE 3. Mean number of errors of each of four groups on final, four-minute sentence test.

0.11, and there were no reliable differences for code errors. The mean performance scores for sentence speed, sentence errors, and code speed are shown in Figures 2, 3, and 4, respectively.

Although certain of the comparisons between pairs of treatments are neither orthogonal nor independent, those with probabilities exceeding 0.05 by Mann-Whitney "U" test (3, pp. 434-435) are shown in Table I.

Perhaps the most concise overview of the results may be seen in Table II, which displays the ranks of the four groups on each of the four criterion variables and the mean rank for each group.

It seems clear from these results that as measured at an early stage of training the initial use of the traditional, single-finger location drill sequences to teach key locations produces speeds which are inferior to those produced by learners trained on word

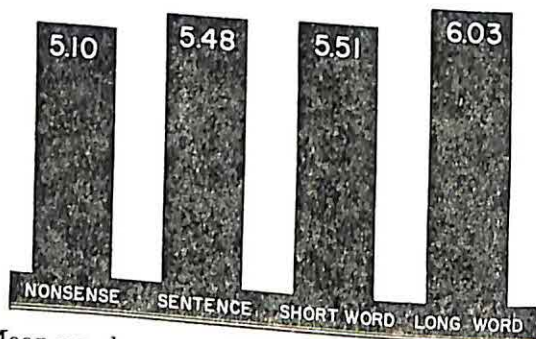


FIGURE 4. Mean words per minute of each of four groups on final, two-minute test on random, five-letter, equal-frequency code groups.

TABLE I.—PROBABILITY OF DIFFERENCE BETWEEN MEANS FOR VARIOUS PAIRS OF GROUPS BY MANN-WHITNEY "U" TEST

Experimental Pair*	Probability for		
	Sentence speed	Sentence errors	Code speed
Sentence-nonsense	0.01	0.02	0.008
Short word-nonsense	0.001		
Long word-nonsense	0.0004		

* The first-named group in each pair is the superior group. Only pairs for which differences are significant at the 0.05 level or better are shown.

TABLE II.—RANK OF MEAN SCORES FOR FOUR CRITERION VARIABLES FOR EACH OF FOUR EXPERIMENTAL GROUPS AND MEAN OF RANKS FOR EACH GROUP

Group	Rank on				Mean Rank
	Sentence		Code		
	Speed	Errors	Speed	Errors	
Nonsense	4	4	4	3	3.75
Sentence	3	2	3	2	2.50
Short word	2	1	2	1	1.50
Long word	1	3	1	4	2.25

or sentence materials. There is also an insufficiently reliable trend toward fewer errors on sentence materials by word and sentence typists. In general, both for sentence and code typing, the use of word or sentence materials appears to give learners an initial advantage over those trained on nonsense sequences.

DISCUSSION

Although this experiment was not designed to test directly possible reasons for the general superiority of word and sentence over nonsense materials, several speculations may be offered. It will be recalled that the stroking habits of typewriting were described as hierarchic, that nonsense materials by their very nature are designed to keep stroking on the lowest or letter-by-letter level. Combinations such as *sz*, *fv*, *jm*, etc., do not exist in the language and it does not seem likely that such combinations

would be handled in the grouped or patterned fashion which characterizes the use of higher-order stroking habits. In word and sentence materials, on the other hand, the learner practices the varied and characteristic sequences of common combinations of letters which the language contains. Such materials create opportunities, especially for the more able learner, to attempt, perhaps, a slight crowding of the more common letter sequences. The chief factor accounting for results is thus thought to be differences in the restrictiveness of the materials in allowing for individual differences in readiness to attempt higher-order stroking habits. A second consideration arises from the presence in nonsense materials of sequences which do not exist in the language. To whatever extent such sequences get stroked as sequences, they may interfere with the subsequent typing of meaningful materials and may have to be "unlearned." Third, the letter-combinations in location drill materials are very much fewer in number than those of dictionary materials. Thus, the word and sentence materials have the well-known advantage of greater variety of responses in the practice and thus greater transfer value than do the less varied responses of the traditional drill materials. Last, intrinsic interest factors are probably to the advantage of meaningful over nonsense materials.

SUMMARY

This experiment evaluated four types of early keyboard learning practice materials in typewriting: single-finger nonsense sequences, short words, long words, and sentences. Each of these four types of materials was used by a different group of learners for the first hour of practice. During the second and last hour of practice, all learners used the same code and alphabetic sentence materials. Statistical analysis of the speed and error scores of these learners on final tests using code and alphabetic sentence materials showed the word and sentence typists to be, in general, superior to those initially trained on traditional drill sequences. The chief factors which may account for the superiority of word and sentence over nonsense materials are thought to be: (a) the greater allowance for individual differences in ability to attempt stroking habits of a higher order, (b) the absence of possible interference from earlier practice on sequences which do not exist in the language, (c) the

greater variety of responses allowed by the practice, and (d) the higher levels of attention, concentration, and interest.

REFERENCES

- 1) W. F. Book, *Learning to Typewrite*. New York, Gregg Publishing Co., 1925.
- 2) W. F. Book, *The Psychology of Skill with Special Reference to Its Acquisition in Typewriting*. New York, Gregg Publishing Co., 1925.
- 3) H. M. Walker and J. Lev, *Statistical Inference*. New York, Henry Holt, 1953.

BEHAVIORAL CONTROL AND INTELLECTUAL ACHIEVEMENT OF SECONDARY SCHOOL BOYS¹

NORMAN TALLENT²

Teachers College, Columbia University

Information on the relationships between the so-called non-cognitive personality variables on the one hand, and intellectual ability and vocational capacity on the other is often pertinent to problems of educational diagnosis and guidance. Factors of behavioral control, in particular, are thought to enter intellectual attainments, a low degree of behavioral control, or impulsiveness, being commonly thought to be incompatible with a high degree of intelligence. Webster's (14) defines "impulsive" as "impetuous," "quick," and "ill-considered." A psychological dictionary (6) apparently reflects the belief of most psychologists in stating that impulsiveness is characterized by a lack of "... foresight or prudence ... in which the cognitive aspect is subordinated to the conative-affective aspects."

Some clinical evidence in support of this view has been accumulated (10), but relationships between behavioral control and intelligence have not been demonstrated in a school situation. Thurstone (11) reported correlation coefficients between a factor labeled "Impulsive" and some measures of intelligence. Correlations of scores on the Impulsive scale of the Thurstone Temperament Schedule with the total score and Q and L scores of the ACE, and with the total score and the subtest scores of the PMA, ranged from -0.073 to 0.152. All but the largest of these correlations, that between the Impulsive score and the PMA Word Fluency score, failed to reach the 0.05 level of significance.

¹ Based on a portion of the thesis submitted in partial fulfillment of the requirements for the Ph.D. degree at Teachers College, Columbia University. Appreciation for assistance is extended to the writer's dissertation committee, Dr. Laurance F. Shaffer, Dr. Robert L. Thorndike, and Dr. Edward Joseph Shoben, Jr.

² Now at Veterans Administration Center, Kecoughtan, Virginia.

DEFINITION OF PROBLEM

The present study was designed to explore the relationship between behavioral control in school and performance on a wide range of intelligence tests. If it is agreed that behavioral control is itself an expression of intelligence, then a positive relationship should exist. Consideration was also given to the possibility that a greater relationship exists between behavioral control and intellectual achievement on some tests than on others. With this orientation, the following hypotheses were formulated:

(1) Subjects who are rated as showing a high degree of behavioral control have significantly higher mean intelligence test scores than subjects who are rated as impulsive.

(2) The above differences are significantly greater on tasks presumed to demand concentration, a high order of conceptualization, or evidence of long-term investment in intellectual matters, such as is seen in the accumulation of knowledge, than on tasks judged not to demand these qualities.

ASSESSMENT OF BEHAVIORAL CONTROL

A continuum of behavioral control was postulated with a high degree of control at one end, an average degree of control in a central position, and impulsiveness at the other. Nine supposed forms of undercontrolled behavior were then gleaned from an authoritative dictionary (14), from extensive psychoanalytic discussions of undercontrol (5, 9), and from an examination of items on the Thurstone Impulsive Scale (12). Together with their opposite polarities, these forms of behavior may be listed as:

- (1) Slowness of action vs. quickness of action.
- (2) Slowness of thought vs. quickness of thought.
- (3) Thought orientation vs. action orientation.
- (4) Patience and persistence vs. impatience and lack of persistence.
- (5) Carefulness and conscientiousness vs. carelessness and lack of conscientiousness.
- (6) Conservative spirit vs. adventurous spirit.
- (7) Socially positive behavior vs. social delinquency.
- (8) Rigidity of procedure and thought vs. lability of procedure and thought.
- (9) Stability of mood vs. lability of mood.

BEHAVIORAL CONTROL RATING SCHEDULE

A provisional Behavioral Control Rating Schedule (BCRS) was constructed, consisting of sixteen rating scales which were based on the above nine continua of behavioral control. Seven of the continua were represented by two scales each; two continua, speed of action and thought, were represented by one scale each in view of the consideration that intelligence might strongly influence speed. The first eight scales were intended to refer to approximately the same aspects of behavior as the last eight scales. Each scale was composed of four descriptive statements, having reference, insofar as possible, to concrete aspects of behavior. The statements of each scale extend from a high degree of behavioral control to undercontrol.

To use the schedule, one statement on each scale is checked. The number of the statement checked, 1, 2, 3 or 4 is taken as an index of control with reference to the aspect of behavior measured by the scale. The sum of the numbers checked on all the scales yields a figure which constitutes an index of general behavioral control.

STATISTICAL QUALITIES OF THE BCRS

To test the statistical qualities of the schedule, ratings were accomplished on eighty-six ninth-grade boys from the Elgin, Illinois, school system, as follows: (1) students and teachers in separate meetings with the experimenter were acquainted with the BCRS and asked to check on a mimeographed class roster the names of the students they felt they could rate on the schedule, (2) each student and each teacher completed ratings on five or fewer students from the names they had checked. In this manner, an average of 4.63 peer ratings and 4.07 teacher ratings were obtained on each student.

The corrected split-half reliability of the schedule, as rated by the students was 0.77. The teachers showed greater consistency in their ratings, a reliability coefficient of 0.90 being found.

The several ratings on each student were averaged and estimates of the between-rater reliabilities of the averaged multiple ratings of the students were computed by procedures based upon analysis of variance (4). The reliability of the averaged ratings was 0.86 for the student raters and 0.92 for the teacher raters. It thus appeared that averaged ratings of the BCRS could reliably be used

to separate students into different degrees of behavioral control as defined by this rating method.

A correlation between the averaged ratings of the students and the averaged ratings of the teachers was also computed. An r of 0.80 was obtained, indicating that these two groups of raters tended to evaluate the behavior of students similarly with the BCRS.

To determine the degree to which the various scales contribute to an estimate of behavioral control, an item analysis was carried out on a sample of sixty schedules completed by the teachers and sixty schedules completed by the students. The two scales which refer to speed of response were negatively correlated with the other postulated aspects of behavioral control, i.e., with the full BCRS, and were eliminated from the schedule. Otherwise, the various scales tend to contribute substantially to the full score. Scale F, as rated by the students, was anomalous in this respect, apparently because of some misleading wording which was retrospectively discovered and altered. The items of the revised BCRS are listed as Table I, the figures in parentheses preceding each scale referring to the eta correlation of that scale with the full score. The top figure is that of the student raters; the bottom figure is that of the teachers.

DESIGN AND CONDUCT OF THE EXPERIMENT

Classification of subjects into experimental groups. The subjects were one hundred and ninety-eight ninth-grade boys of the Elgin, Illinois, public school system. The degree of behavioral control of these subjects was appraised in terms of the averaged BCRS ratings by their teachers. An attempt was made to secure five ratings for each subject, but practical consideration reduced the number obtained to an average of 4.33. From the one hundred and ninety-eight rated, eighty were selected on the basis of the mean BCRS scores assigned them and placed in two groups. The forty subjects who received the highest averaged rating scores were designated the "Impulsive Group" (I Group) and the forty subjects who received the lowest averaged rating scores were labeled the "Controlled Group" (C Group). The mean rating scores for the I and C groups respectively were 41.73 and 25.05. It is noted that "impulsive" refers to the most poorly controlled subjects in a sample of normal school boys and the findings on

this group cannot be applied to persons having various degrees of pathological impulsiveness.

The intelligence test battery. The battery was selected for the purpose of testing the hypotheses of the study on a variety of intellectual tasks which represent current techniques for measuring intelligence. The battery included group tests and individual tests, verbal tests and performance tests, tests based on the factorial isolation of mental abilities, and tests which purport to measure intelligence without the intrusion of cultural influences. A number of orientations to intelligence influenced the design of the various tests. The tests were:

- (1) Wechsler-Bellevue Intelligence Scale, Form I (15).
- (2) Porteus Mazes, Vineland Edition (1933) (8).
- (3) American Council on Education Psychological Examination for High School Students, 1948 Edition (ACE) (1).
- (4) SRA Primary Mental Abilities for ages 11-17 (PMA) (13).
- (5) A Culture-Free Test (2).

Classification of tests. All of the intelligence tests used in the study were examined by the experimenter in terms of the considerations which led to the formulation of Hypothesis 2. Tests which were judged to put impulsive subjects at a disadvantage, and hence show a high positive relationship between behavioral control and intelligence, were labeled "Relationship" or R tests. Tests which were judged not to put impulsive subjects at a disadvantage, and hence might show little or no relationship between intelligence and behavioral control, were labeled "No Relationship" or N tests.

All but seven tests were thus hypothesized to show a high relationship. Five of these were the performance tests of the Wechsler scale. The anticipated quickness in making decisions by the impulsive subjects, along with the ready opportunity afforded by the Wechsler performance tasks for the spontaneous correction of errors, were considered to be factors which would cause the scores of impulsive subjects to approximate those of other subjects. The Digit Span test was not expected to show relationship because of a clinical observation (9) that success on the test demands essentially a passive attitude. It was expected that impulsive subjects would also be enabled to score at approximately the level of their better controlled peers on the Word Fluency test. The underlying

TABLE I.—SCALES OF THE REVISED BEHAVIORAL
CONTROL RATING SCHEDULE

-
- A. — 1. Thinks very carefully before doing anything.
 — 2. Gives at least some thought to a matter before doing something about it.
 (0.78)
 (0.79) — 3. Tends to do something about a matter before he has a chance to give it much thought.
 — 4. Acts before he has a chance to think at all about what he does.
- B. — 1. Sticks to a task no matter how difficult.
 (0.60) — 2. Gives up or quits only on very difficult tasks.
 (0.77) — 3. Gives up or quits difficult tasks fairly quickly.
 — 4. Gives up or quits a task as soon as he runs into any difficulty.
- C. — 1. Does very careful, accurate work.
 (0.63) — 2. Work is fairly careful and accurate.
 (0.82) — 3. Tends to be careless and inaccurate.
 — 4. Does very careless, "sloppy," inaccurate work.
- D. — 1. Will never take a chance.
 (0.79) — 2. Might take a chance.
 (0.81) — 3. Tends to take chances.
 — 4. Takes reckless chances.
- E. — 1. Is exceptionally well-behaved—never gets into trouble.
 (0.84) — 2. Is fairly well-behaved—might get into trouble.
 (0.87) — 3. Gets into trouble quite often—not very well-behaved.
 — 4. Is always getting himself into trouble—very badly behaved.
- F. — 1. Can't be talked into changing his mind.
 (-0.36) — 2. Will change his mind if given good reasons.
 (0.63) — 3. Is quick to take on the ideas of others and change his mind.
 — 4. Immediately takes on everyone else's ideas—doesn't have a mind of his own.
- G. — 1. Does not seem to show his feelings (anger, sadness, happiness, etc.) at all.
 (0.58) — 2. Does not show his feelings very easily.
 (0.50) — 3. Shows his feelings fairly easily.
 — 4. Shows his feelings very easily.
- H. — 1. Spends his free time in serious conversation or study—definitely doesn't care to take part in sports or active games.
-

TABLE I—*Continued*

-
- (0.44) — 2. Prefers serious conversation or study to taking part in sports or active games.
- (0.79) — 3. Prefers taking part in sports or active games to serious conversation or study.
- 4. Spends his free time in taking part in sports or active games—doesn't care at all for serious conversation or study.
- I. — 1. Always shows much patience.
- (0.42) — 2. Generally shows a fair amount of patience.
- (0.81) — 3. Tends to be rather impatient.
- 4. Is always very impatient.
- J. — 1. Is very responsible—can be trusted to do anything.
- (0.77) — 2. Is fairly responsible—can generally be trusted.
- (0.82) — 3. Tends to be irresponsible—probably wouldn't care to trust him.
- 4. Is very irresponsible—can't be trusted to do anything.
- K. — 1. Keeps away from action and excitement.
- (0.49) — 2. Tends to keep away from action and excitement.
- (0.78) — 3. Tends to be where there is action and excitement.
- 4. Is always in the middle of action and excitement—will find him where something is going on.
- L. — 1. Always obeys the rules.
- (0.91) — 2. Usually obeys the rules.
- (0.86) — 3. Usually does not obey the rules.
- 4. Doesn't obey the rules at all—acts as if they don't apply to him.
- M. — 1. Finds it very hard to change his ways of doing things—does things his own way or not at all.
- (0.15) — 2. Can be talked into changing his ways of doing things.
- (0.52) — 3. Readily agrees to another's way of doing things.
- 4. Would just as soon do something someone else's way as his own way.
- N. — 1. Is a very serious person.
- (0.71) — 2. Is a rather serious person.
- (0.79) — 3. Tends to be carefree and gay.
- 4. Is very carefree and gay—really lets himself go and have a good time.
-

rationale was the expectation that impulsive subjects would make rapid, if not critical, associations, whereas it was thought that better controlled subjects would achieve their scores as a result of having superior verbal skills. The low positive correlation found by Thurstone (11) between Impulsive scores and Word Fluency scores further supported the decision to classify Word Fluency as an N test.

Testing program. The three group tests of intelligence (PMA, ACE, Cattell Culture-Free) were administered to one hundred and ninety-four students on whom ratings were available in separate testing sessions. The two individual tests of intelligence (Wechsler-Bellevue I, Porteus Mazes) were administered to all of the subjects comprising the C and I groups.

FINDINGS AND IMPLICATIONS

The data³ indicate that there was a high relationship between intelligence scores and behavioral control ratings. There was a strong tendency for the subjects rated as showing a high degree of control to obtain the highest intelligence scores, with subjects rated as inferior on control achieving the lowest scores. Thus, the C Group means were higher than those of the I Group on twenty-five of twenty-eight subtests at or beyond the 0.01 level of confidence and on two of twenty-eight subtests at or beyond the 0.05 level.⁴ Correlations between behavioral control scores and intelligence also show a pervasive relationship between these variables (Table II). A high score on the BCRS represents low control; therefore the negative correlations between control scores and intelligence scores indicate that the well-controlled subjects tended to obtain the highest intelligence scores and that the impulsive subjects tended to achieve the lowest scores. Hypothesis 1 is therefore clearly supported.

An obvious question arises in connection with these findings.

³ To save printing costs, a table giving the mean scores and differences between means (together with t values for these differences) for the two groups on the tests of the intelligence battery has been deposited with the ADI Auxiliary Publications Project. Order Document No. 5099 from: Chief, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting \$1.25 for photoprints, or \$1.25 for 35 mm. microfilm.

⁴ A group of boys rated average in behavioral control was also tested. The C Group means were significantly higher than those of the Average Group on most of the subtests and the Average Group tended to score higher than the I Group.

TABLE II.—CORRELATIONS BETWEEN SCALES OF
THE BEHAVIORAL CONTROL RATING SCHEDULE
AND GROUP TESTS OF INTELLIGENCE
(N = 194)

Test	BCRS Scale							
	Total	A	B	C	D	E	F	G
ACE	-0.45	-0.46	-0.51	-0.47	-0.34	-0.29	-0.33	-0.12
PMA	-0.37	-0.43	-0.52	-0.49	-0.27	-0.21	-0.33	-0.19
Culture-Free	-0.52	-0.50	-0.54	-0.58	-0.41	-0.45	-0.30	-0.19
	H	I	J	K	L	M	N	
ACE	-0.47	-0.24	-0.38	-0.24	-0.37	-0.26	-0.17	
PMA	-0.29	-0.15	-0.33	-0.10	-0.26	-0.28	-0.18	
Culture-Free	-0.42	-0.34	-0.45	-0.23	-0.47	-0.16	-0.36	

Could the teachers' ratings have been influenced by their knowledge of the students' scholarship, overestimating the behavioral control of the students who achieve academic success and underestimating the control of students who do not? The data of Table II can be interpreted to suggest that some errors were obtained from this source. The rating scales having the highest correlations with the group tests appear logically to concern forms of behavior which reflect affective personality components. Thus, the average correlations of Scales B, C, A and H with the group tests ranged from 0.39 to 0.52. These scales concern, respectively, ability to persevere at a task, carefulness and accuracy of work, tendency to think before acting, and relative preference for serious conversation or study to sports or active games. The average correlations of Scales G, K, M and N with the group tests ranged from 0.17 to 0.24. These scales concern, respectively, control over expression of feelings, tendency to be where there is action and excitement, readiness with which preferred modes of behavior are surrendered, and seriousness of mood. It is to be noted, however, that most of these low correlations are statistically significant. For the size of the sample employed, a correlation of 0.14 is significant at the 0.05 level and a correlation of 0.18 is significant at the 0.01 level. The correlation of 0.80 between teacher ratings of students and student ratings of their peers suggests further that the obtained

results are only partially based on teacher halo effect. The students showed good agreement with their teachers concerning the behavioral control of their fellow students even though it is unlikely that the students would be as cognizant of, or as concerned with, the scholarship of most of their peers as were the teachers.

The second hypothesis of this study was that the differences in mean scores between the C and I groups would be significantly greater on tasks judged to demand concentration, a high order of conceptualization, or evidence of long term investment in intellectual matters, such as is seen in the accumulation of knowledge (R tests), than on tasks judged not to demand these qualities (N tests). Such a finding would indicate that the R tests have a higher relationship to behavioral control than do the N tests.

This hypothesis was tested by the Mann-Whitney U test (7) which was used to estimate the probability that the t values for the differences between R test means and the t values for the differences between N test means were samples from a common population. The t values for the differences in means between the C and I groups were ranked from smallest to largest. The sum of ranks of the t values for the differences between the R test means was found to be proportionately greater than the sum of ranks of the t values for the differences between N test means. The obtained distribution of t values was found to be such as could be expected to occur by chance three times in a thousand if these t values were samples from a common population. Therefore, it was concluded that there were greater differences between the means of these two groups on R tests than on N tests.

To better define the patterns of intellectual functioning, the relative strengths and weaknesses among the two groups were analyzed. An additional advantage to such analysis is that it shows the relationship between behavioral control and intellectual achievement in the various areas measured by the battery independently of general intelligence levels, hence uncontaminated by possible rater bias which stems from knowledge of the subjects' school achievement. This operation was accomplished as follows: (a) subtest scores of the group tests were converted to standard scores (this operation was not necessary for the Wechsler scale because weighted Wechsler subtest scores are standard scores); (b) the deviations of each individual's subtest scores from the average of all of his subtest scores of each scale were computed;

TABLE III.—MEAN DEVIATIONS OF SUBJECTS' SUBTEST SCORES
FROM THEIR AVERAGE SUBTEST SCORES AND
DIFFERENCES BETWEEN THESE DEVIATIONS

Test	\bar{D}_C	\bar{D}_I	$\bar{D}_C - \bar{D}_I$
W-B Arithmetic	-0.05	-2.22	2.17**
W-B Similarities	0.77	-0.17	0.94*
W-B Picture Completion	0.62	1.44	-0.82*
W-B Digit Symbol	-1.15	0.29	-1.44**
PMA Space	-1.71	2.65	-4.36*
Cattell Series	2.94	-0.06	3.00*

* Significant at or beyond the 0.05 level.

** Significant at or beyond the 0.01 level.

(c) the deviations on each subtest for all subjects in each group were added algebraically, and the mean deviations for each group were computed; and, (d) the differences in mean deviations between groups were computed and tested for significance (t). Table III shows the mean deviations and the differences between deviations for all subtests which showed significant differences (0.01 or 0.05) when the mean deviations between groups were compared.

The number of tests on which a significant relationship was found between behavioral control and mean deviations was greater than might be expected on a chance basis if there were no real differences between groups in mental functioning. In considering the practical significance of these findings, it is noted that most of the tests are factorially complex when administered to school children (3). Hence, for the most part, the relationships between behavioral control and intellectual achievement demonstrated here are not relationships among neatly isolated psychological traits. Similarly, we may be quite sure that a number of factors contribute to achievement in any given school subject. The educational implications of the obtained relationships between behavioral control and intellectual achievement might therefore be examined to determine if any descriptive similarities may be found between performance on the tests in Table III and functioning in various school subjects or curriculum areas.

The Controlled Group shows superiority on the Wechsler Arithmetic test, a finding which is consistent with Rapaport's hypothesis that the ability to concentrate contributes to success on this test (9). Since the basic operations required on the Arithmetic test

enter the various quantitative skills, especially mathematics and science, the implications seem clear. The Controlled Group is also strong on the Similarities test and on the Series test, both tasks appearing to require considerable reasoning, including the use of abstract processes.

The Impulsive Group is relatively proficient on the Picture Completion test, a task which seems to call for an awareness of the environment, the ability and willingness to report evidences of such awareness, and very little intervening mental manipulation of data. These are minimal requisites for functioning in all school subjects. Of the more positive skills, the relative superiority of the Impulsive Group on the Space test indicates some ability to imagine the positions of objects in space. Findings on the Digit Symbol test suggest that, relative to intelligence, the Impulsive boys are superior on tasks of sensorimotor speed and coordination. There is reason to infer, therefore, that behaviorally well-controlled boys function at their best in areas demanding quantitative and abstract reasoning—areas in which more poorly controlled boys are relatively handicapped—and that these more impulsive school boys have at least some of the intellectual abilities which enter various forms of mechanical work and possibly also such tasks as engineering drawing and blueprint reading.

The design of the experiment makes it likely that the number of relations found here between behavioral control and intellectual achievement, when data are analyzed independently of general intelligence level, is minimal. It is possible that some real relationships were obscured by the complexity of factors entering performance on any given task, with competing tendencies giving rise to scores which fail to show significant differences when the C and I groups are compared. For example, the data analysis did not support the belief that poorly controlled boys are superior in Word Fluency. It is possible, however, that rapid, uncritical associations contributed to the scores of those low in control, this advantage being offset by the superior vocabulary of the well controlled.

The obtained relationships can not, of course, demonstrate causality or direction. It appears plausible that impulsive modes of behavior can detract from intellectual achievement, but it is also possible that, at least in some cases, behavioral undercontrol is based on lack of adequate school achievement—an observation

of many teachers. Where emotional factors are primary in detracting from intellectual achievement, counseling might be of maximum value in releasing potential. Where the student misbehaves due to frustration over nonachievement, course guidance might be most indicated.

SUMMARY

Relationships between the behavioral control and intellectual achievement of secondary school boys were investigated in view of the pertinence of such relationships to problems of educational diagnosis and guidance. Behavioral control was defined in terms of ratings by teachers on the Behavioral Control Rating Schedule, an instrument developed for this study. Intellectual achievement was determined by administering an extensive battery of intelligence tests. The major finding was that a pervasive relationship exists between behavioral control and intelligence, those boys rated as highest in control having the highest intelligence. The degree of relationship, however, was significantly higher on some tests than on others. Certain relative strengths and weaknesses were also found to be associated with different levels of behavioral control. Some of the possible implications of these findings for student guidance were discussed.

REFERENCES

- 1) *American Council on Education Psychological Examination for High School Students*, 1948 edition. Princeton, N. J., Educational Testing Service, 1948.
- 2) R. B. Cattell, *A Culture-Free Test*. New York, The Psychological Corp., 1944.
- 3) P. C. Davis, "A Factor Analysis of the Wechsler-Bellevue Intelligence Scale, Form I, in a Matrix with Reference Variables," *American Psychologist*, (Abstract). 7: 296-297, 1952.
- 4) R. L. Ebel, "Estimation of the Reliability of Ratings," *Psychometrika*, 16: 407-424, 1951.
- 5) O. Fenichel, *The Psychoanalytic Theory of Neurosis*. New York, Norton, 1945.
- 6) P. L. Harriman, *The New Dictionary of Psychology*. New York, Philosophical Library, 1947.
- 7) H. B. Mann and D. R. Whitney, "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics*, 18: 50-60, 1947.
- 8) S. D. Porteus, *Porteus Mazes, Vineland Edition (1933)*. New York, The Psychological Corp., 1933.

- 9) D. Rapaport, M. Gill and R. Schafer, *Diagnostic Psychological Testing*, 2 volumes. Chicago, Year Book Publishers, 1945.
- 10) R. Schafer, *The Clinical Application of Psychological Tests*. New York, International Universities Press, 1948.
- 11) L. L. Thurstone, *Examiner Manual for the Thurstone Temperament Schedule*. Chicago, Science Research Associates, 1950.
- 12) L. L. Thurstone, *Thurstone Temperament Schedule*. Chicago, Science Research Associates, 1949.
- 13) L. L. Thurstone and Thelma G. Thurstone, *SRA Primary Mental Abilities*. Chicago, Science Research Associates, 1947.
- 14) *Webster's New International Dictionary of the English Language*, second edition, unabridged. Springfield, Mass., G. & C. Merriam, 1934.
- 15) D. Wechsler, *Wechsler-Bellevue Intelligence Scale (Form I)*. New York, The Psychological Corp., 1947.

BOOK REVIEW

V. J. MCGILL. *Emotions and Reason*. Springfield, Ill.: Charles C Thomas. 1954. pp. 122.

When a man loses his head, he acts stupidly, his conduct is described as emotional. When a man thinks clearly and behaves in a deliberate and rational fashion, it is assumed that emotion is absent or suppressed or under control. This the author does not agree with—hence the book on *Emotions and Reason* to prove the need for applying what he considers to be a more modern approach—elimination of the sharp division between emotion and reason.

His approach, the author believes, will eliminate the split between theory and practice. In this respect, he quotes John Dewey as saying that this split "is probably one of the chief sources of the maladjustments and unendurable strains from which the world is suffering." The author ends the book with an expression of hope that this short book of his will help in the general trend that he sees developing of the functional theory of emotions that will help the opponents of dualism.

Generally speaking, the book is expressed in rather abstract language throughout. In his attempt to summarize the present literature on emotions or reason, the author depends largely on philosophical background for interpreting and discourse. The argument is presented under three chapters: Needs, Emotions and Ethics, and the Emotions. For his background and for the words he uses in discussing the problems, the author gives credit to and shows familiarity with the writings of his former teachers, Edmund Husserl, G. E. Moore and Ralph Barton Perry. Psychologists he particularly likes to quote are: E. C. Tolman and D. O. Hebb.

H. MELTZER

Psychological Service Center, St. Louis, Mo.

